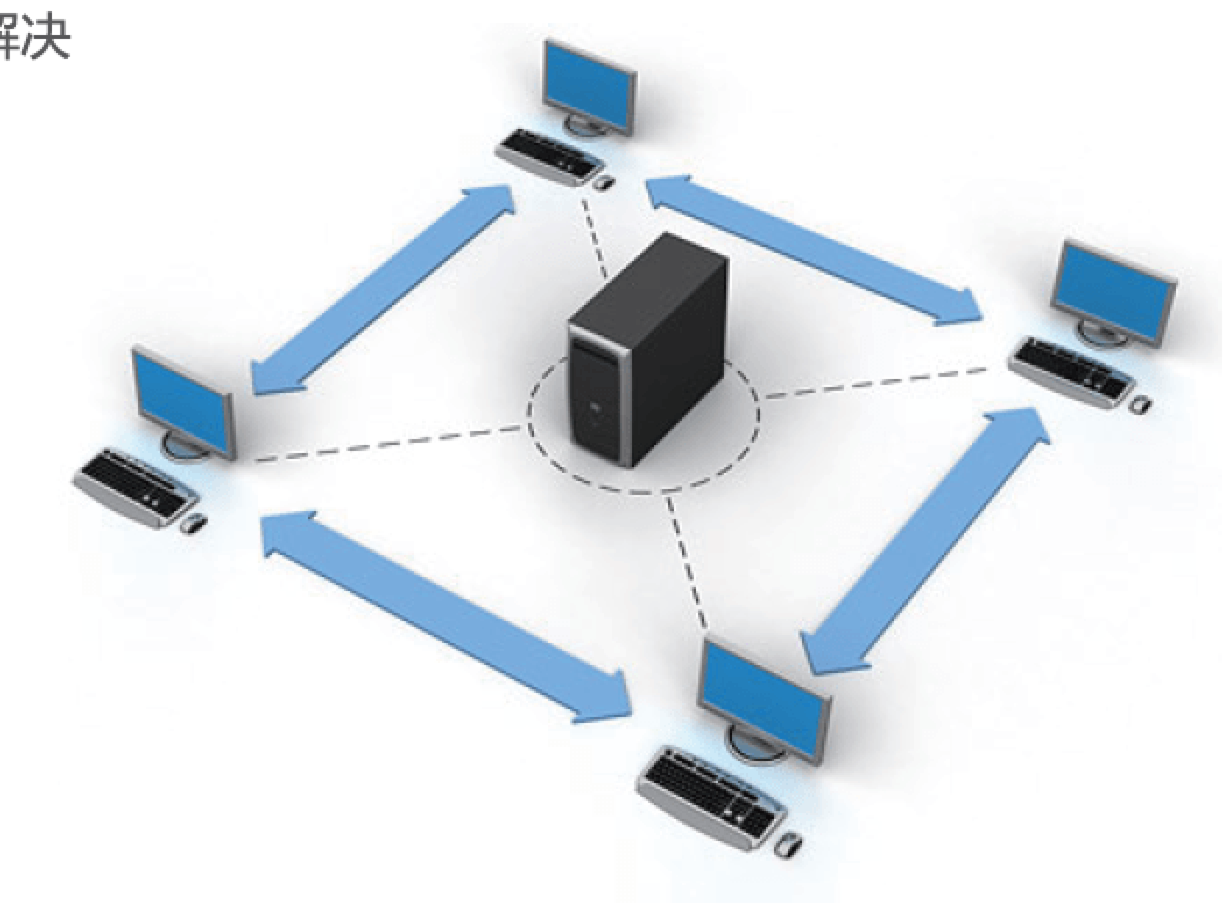
为海量视频处理
提供解决方案

Visual Big Data Basis and Applications

视觉大数据
基础与应用

谢剑斌　等编著

- 结合海量视频分析与搜索在相关行业的应用，解决视频大数据处理中的诸多核心问题
- 详细阐述人脸搜索系统、车牌搜索系统、车标搜索系统、暴力行为检测系统、可疑行为检测系统、海量视频摘要系统和海量视频管控平台等实际案例

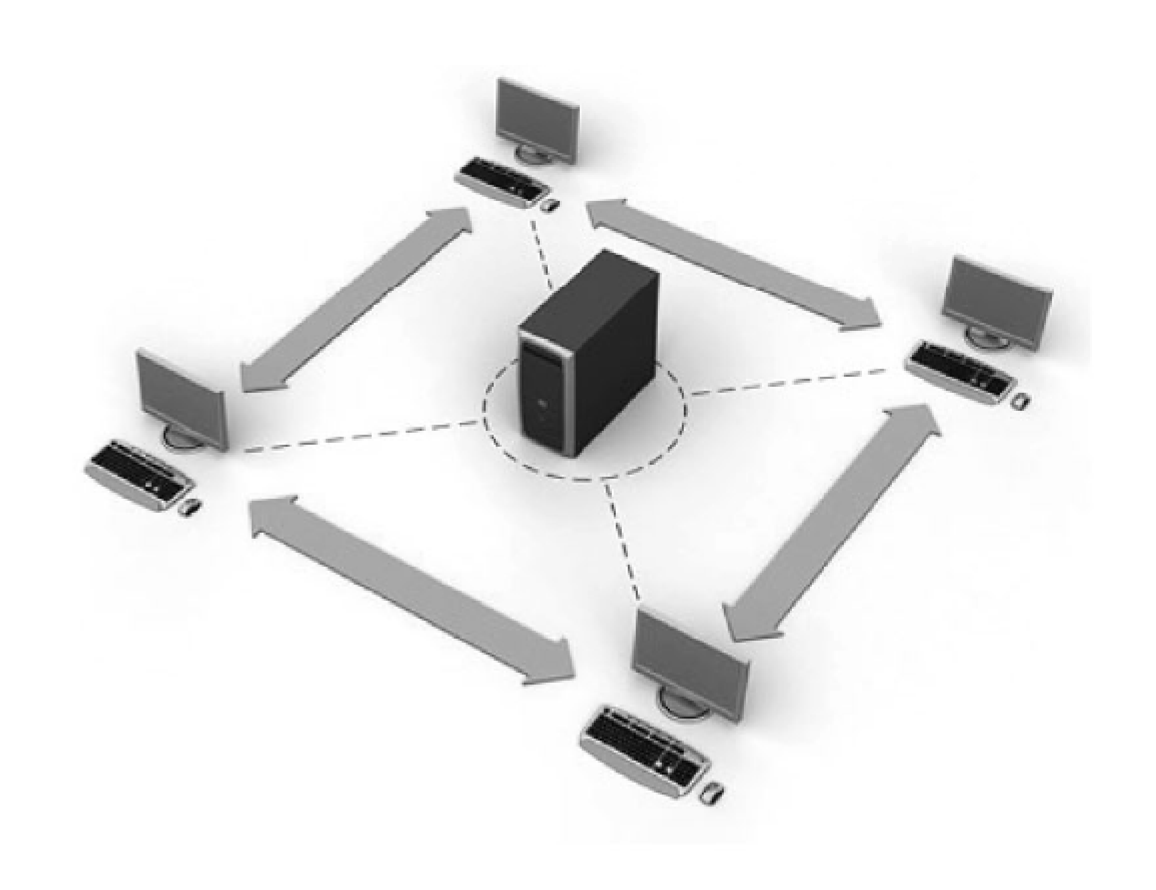


清华大学出版社

视觉大数据

基础与应用

谢剑斌 刘 通 闫 玮
李沛秦 王 勇 谭 筠 编著



清华大学出版社
北 京

内 容 简 介

本书是视频大数据处理领域的著作。为使读者全面了解海量视频分析与搜索的基础知识及应用方法，本书首先介绍海量视频概论、海量视频模型、海量视频管理和海量视频分析等相关基础知识，然后具体阐述面向大数据的大规模人脸搜索系统、面向高清卡口的车辆车牌与车标等信息搜索系统、暴力行为检测系统、可疑行为检测系统、海量视频摘要系统和海量视频管控平台等典型的海量视频分析与搜索实例，并将海量视频分析与搜索领域的新技术和新成果贯穿于全文的描述之中。

本书主要适用于从事海量视频分析与处理领域的应用开发和工程施工技术人员阅读。

图书在版编目（CIP）数据

视觉大数据基础与应用/谢剑斌等编著. —北京：清华大学出版社，2015

ISBN 978-7-302-39122-7

Ⅰ.①视… Ⅱ.①谢… Ⅲ.①视频编辑软件 Ⅳ.①TN94

中国版本图书馆 CIP 数据核字（2015）第 012817 号

责任编辑：王金柱
封面设计：王　翔
责任校对：闫秀华
责任印制：
出版发行：清华大学出版社
　　网　　址：http://www.tup.com.cn，http://www.wqbook.com
　　地　　址：北京清华大学学研大厦 A 座　　　　**邮　　编：**100084
　　社 总 机：010-62770175　　　　**邮　　购：**010-62786544
　　投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
　　质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：
经　　销：全国新华书店
开　　本：180mm×230mm　　**印　　张：**16　　**字　　数：**393 千字
版　　次：2015 年 3 月第 1 版　　**印　　次：**2015 年 3 月第 1 次印刷
印　　数：1~3000
定　　价：49.00 元

产品编号：056802-01

前 言

常言道“百闻不如一见”，人类感官接受的各种信息约有 80%来自视觉。视频和图像等可视化信息是对客观事物形象、生动的描述，是直观而具体的信息表达形式，是人类社会最重要的信息载体。随着光学成像、数字视频、计算机、信号处理等技术的快速发展，以及人类社会对信息获取、安全保卫、智能服务等应用的迫切需求，视频与图像日益受到人们的青睐。海量视频处理在视频图像与内容描述之间建立映射关系，通过视频图像分析来理解场景内容，如人脸识别、行为识别、车牌搜索、车标搜索、视频摘要等。

本书是作者十多年研究海量视频处理的心血结晶，可作为信息、计算机、自动化、电子与通信等学科专业高年级本科生和研究生的实践教材，也可以作为从事海量视频处理领域技术人员的参考设计资料。全书分为 10 章，首先详尽地介绍海量视频的模型、管理、分析的基本理论；然后深入地阐述大规模人脸搜索、高清卡口车辆信息搜索、暴力行为检测、可疑行为检测、海量视频摘要等系统的实施方案和实验仿真，并提供配套的源代码和视频库；最后以某个市级公安局应用为参考，深入浅出地描述海量视频管控平台。

本书由国防科技大学电子科学与工程学院数字视频课题组编著，谢剑斌教授主编，第 4 章、第 10 章由谢剑斌执笔，第 7 章、第 8 章由刘通执笔，第 1 章、第 9 章由闫玮执笔，第 2 章由谭筠执笔，第 3 章、第 6 章由李沛秦执笔，第 5 章由王勇执笔，全书由谢剑斌统稿、修改和完善。在编著过程中得到国防科技大学庄钊文教授、唐朝京教授的关心和指导，林成龙、刘双亚、穆春迪、许浩、戴超、李润华等为本书的编著做了大量工作，国家自然科学基金项目（61303188）对本书的相关研究工作进行了资助，在此一并致谢！由于时间有限，书中若有纰漏之处，敬请广大读者批评指正。

配套资源的网站为：www.kedachang.com。

编　者

目 录

第 1 章

海量视频概述

2008 年 9 月，*Nature* 推出封面专栏“大数据（Big Data）”，阐述大数据在数学/物理/生物/信息等基础学科、工程技术、社会经济领域中扮演着非常重要的角色。随后 *Science*、《华尔街日报》、《求是》等权威媒体大篇幅介绍大数据，大数据在 Google、百度、必应等成为搜索热点。大数据成为当代的标志符号，海量视频是大数据的重要形态，即视觉大数据。

1.1　视觉大数据

如图 1.1 所示，符合 4 个 V 的数据为“大数据”，海量视频就是视觉大数据。

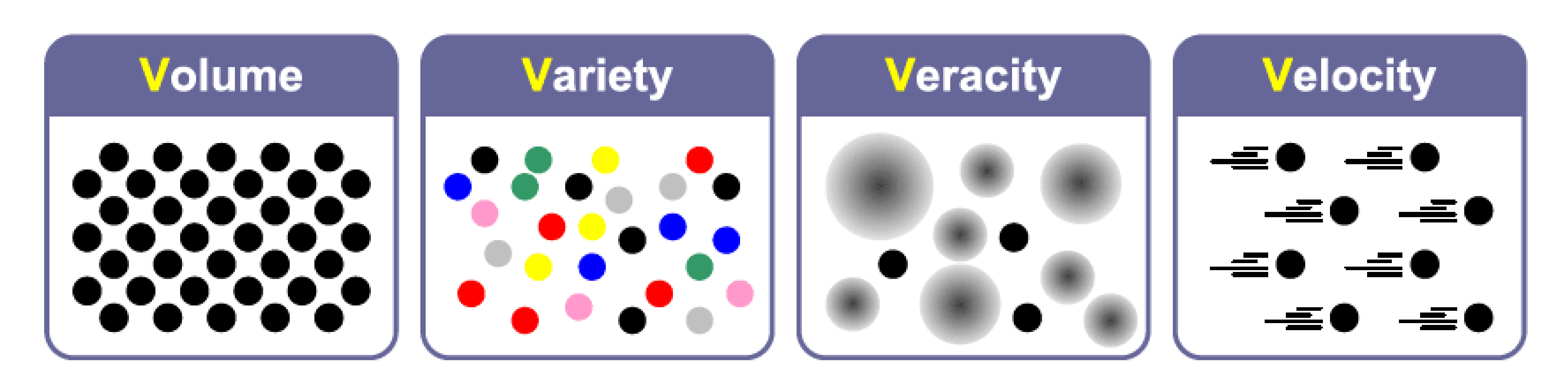


图 1.1　大数据的 4 个 V

1. Volume

Volume（大量）指数据总量巨大。

以视频监视系统产生的视频数据为例，近年来随着各个城市联网视频监控系统以及高清摄像头的普及，视频数据快速增长。以某个部署 10 000 个标清摄像机的中等城市为例，每个摄像机每秒采集到的视频数据经压缩编码后的数据量约为 720（画面宽度）×576（画面高度）×2B（使用 4:2:2 采样）×25（帧率）×0.01（H.264 编码平均压缩比）=207360B≈0.2×10^6 B（0.2MB），每天产生的视频数据量约为 0.2×10^6 B×60（秒）×60（分）×24（小时）×10000（台）=172.8×10^{12} B（172.8TB），每个月产生的视频数据量约为：172.8×10^{12}B×30（天）=5.184×10^{15}B（5.184PB）。

在实际系统中，为了降低存储压力，通常仅存储关键事件（如人、车、物）的画面和描述信息，保守估计平均每个摄像机每 10 秒发生 1 个事件（即每秒 0.1 个事件），则每天产生的事件记录约为 0.1（条/秒）×60（秒）×60（分）×24（小时）×10000（台）=86.4×10^6 条，每年产生的事件记录约为 86.4×10^6 条×365（天）≈31.5×10^9 条；假设记录需保存 3 年，每条记录平均需要占用 0.4MB（2 秒视频）的存储空间，则所需的总存储空间约为 31.5×10^9 条×3（年）×0.4MB≈37.8×10^{15}B（37.8PB）。

视频分享网站产生的视频数据量同样巨大，据统计，2012 年 YouTube 网站上每分钟由用户上传的视频数据平均超过 40 小时，按标清视频数据量计算，每年产生视频数据量约为 0.2×10^6（字节/秒）×60（秒）×60（分）×40（小时）×60（分）×24（小时）×365（天）≈15.14×10^{15}B（15.14PB）。近几年随着智能手机等具备视频采集功能设备的普及，视频上传量更是呈现爆发式增长。

2. Variety

Variety（多样）指数据种类很多。

如表 1.1 所示，海量视频数据的来源多种多样、内容包罗万象。

表 1.1　海量视频数据分类

分类依据	具体类别
信号形式	模拟视频、数字视频
分辨率	CIF、4CIF、D1、720P、1080I/P、2K、4K 等
色彩	真彩色、灰度、伪彩色（如热像仪成像）等
来源	监控系统、影视作品、个人拍摄等
场景	室内、室外，白天、夜晚
环境	晴天、阴雨、大雾等
摄像机姿态	固定式、运动式
编码格式	MPEG-1/2/4、H.261/263/264/265、AVS、SVAC 等

3. Veracity

Veracity（精确）指数据的数据总量大、价值密度低。

该特点在海量视频数据上尤为突出。以监视视频为例，在 1GB 的监视视频中，有用的数据总量可能仅仅只有 10MB。

4. Velocity

Velocity（速率）指数据的流通速度快、实时性强。

监控视频数据放映的是监控场景的实时情况，具有实时性；在对监视视频进行处理时，处理速度越快，实时性越高，数据所体现出的价值就越大。

1.2　关键技术

海量视频数据是由传统的分立多源视频数据形成的聚合体，不仅包含原始视频数据的全部数据量，而且通过分析多源视频的内在联系，还可以挖掘出单个视频数据无法提供的信息，实现 1+1>2 的超越。

下面介绍与海量视频相关的关键技术。

1. 存储与管理

海量视频数据集记录数众多，容量巨大，导致在采集、传输、存储、处理、检索、共享、分析、显示数据集时产生巨大障碍，无法采用传统的基于单机或小规模服务器集群的数据库、文件存储、分布式处理技术，必须采用基于大规模计算集群或数据中心的可灵活扩展、可容错、大规模分布式并行处理技术。“云计算”（大规模集群分布式并行计算技术）被认为是当前的最佳解决方案，已经成为智慧城市物联网应用的组成部分。

2. 分析与识别

传统的视频监控系统依赖人工对监控视频进行实时查看和后期搜索。由于人眼并非可靠的观察者，人工值守容易忽视画面监控造成失误；同时人工搜索的效率异常低下，不能满足海量视频的应用需求。

视频分析与识别技术是解决该问题的关键。譬如：在前端高清网络摄像机中植入智能功能，通过视频分析，对高清监控场景中的人或物进行分析和识别，对异常现象产生提示或报警；通过网络将前端视频数据汇总到中心服务器，借助高性能硬件进行实时分析和识别；使用分布式计算和云计算技术，挖掘历史记录视频中的有用信息。

3. 摘要与搜索

海量视频数据的价值密度低，完整存储时会浪费巨大的存储空间。利用视频图像处理（如视频浓缩、摘要、增强等）、模式识别、海量数据分类存储以及视频搜索等技术，对海量的存储录像等原始信息进行分析和挖掘，对于目标的特征、行为、联接关系等信息内容，形成各种分类的特征信息库、元数据和索引等，提供统一接口供外部应用搜索，通过有限的线索，达到快速关联和可靠定位的功能。

1.3 应用领域

下面介绍海量视频的典型应用。

1. 情报侦察领域

在公开的媒体视频数据中有时会包含某个重要目标的局部特征片段，情报机构通过分析海量视频数据，将包含类似目标的视频数据进行提取和汇总，有可能挖掘出有价值的目标信息。

2. 公共安全领域

通过分析遍布大街小巷、车站码头、商场酒店等场所的摄像机数据，借助视频分析技术，安全部门可以及时发现异常情况，并在第一时间做出响应，搜寻事发现场的可疑目标及其去向；借助人脸搜索技术，通过和公安系统嫌疑人信息数据库对接，可以及时发现网上追逃的嫌疑人员等。

3. 智能交通领域

通过分析管辖范围内所有道路摄像机的监视数据，实时分析道路交通流量，交通主管部门可以综合分析和统计全城的交通状况；通过建立统一的车辆信息数据库，借助车牌识别、车型识别、车标识别技术，快速发现套牌车和假牌车，快速搜索并定位特定车辆的轨迹和位置。

4. 休闲娱乐领域

网络视频点播已经成为广播电视传播的重要方式，通过建立分布式云存储架构，用户在任何时间、任何地点，只要通过联网终端，就可以随时点播和观看喜欢的视频节目，以便更好地安排工作和休闲时间。

5. 个性广告领域

网络广告已经成为广告业的重要分支，从业者通过收集、分析用户与广告间的海量互动视频，可以分析出什么内容的广告更能吸引客户、什么长度的广告不会引起用户的反感、什么时段适合哪些广告的投放、什么网站的用户更倾向于哪些类型的广告等。

1.4　挑战与发展

1. 面临的挑战

海量视频数据具有庞大的数据量和信息量，相关领域的深入研究和有效应用面临巨大挑战。

（1）高效存储

海量视频数据对传输、存储和计算的带宽要求巨大，由于海量视频数据量的急速扩大，大规模计算需求越来越多，处理技术尚未取得重大突破，堆砌高配硬件成为唯一选择，导致硬件投资增长迅猛。如何有效利用已有硬件、避免重复建设、改进硬件性能、降低系统成本是当前面临的重大问题。

摄像机每天 24 小时不停地工作，如实记录发生的一切，而对于用户来说可能大部分信息无效。为了提高海量视频数据的信息密度，亟需视频摘要与搜索。

（2）快速分析

原始视频数据是非文本、非结构化的数据，对视频内容进行建模和数学表述决定提取性能。如何根据实际需求选用合适模型、如何优化已有模型以满足特定需求、如何衡量模型的表述性能是应用要面临的复杂问题。

在视频监控业务中，错看、漏看、来不及看等是常见困扰。海量视频数据的回溯给安全人员带来生理与心理的双重挑战，经常有看到吐、看到晕等无奈情况。

视频分析的效率决定价值，更低的延迟、更准确的分析是智慧城市的普遍需求。现有技术对 TB 级的数据进行分析和检索需要花费数小时的计算，不能胜任时效性需求。要深入研究适用于海量视频数据实时分析与识别的先进算法和计算模型，实现海量视频数据的模糊查询、快速检索和精准定位。

（3）优化应用

在视频监控业务中，看只是信息采集方式之一，用才是业务拓展的根本。视频监控业务的效率问题成为阻碍产业发展的关键瓶颈。

随着摄像机覆盖广度、密度的增大，视频数据量呈指数级上升，而视频监控数据的

使用效率却在下降，大量的视频数据仍然是独立的、零散的，散布在各个行业与单位独立的系统中，没有联网共享。

在视频监控业务网络化之后，网络设备越来越多，但是设备利用率相对较低，很多计算资源处于闲置状态，没有实现资源的最大化利用，运算效率很低。

2. 发展方向

（1）分布式存储

如果类比水库蓄水方式，典型的网络视频监控数据存储模型是一个由小溪汇聚河流、再汇聚到水库的蓄水方式。小溪数量增多、水量增大是水库蓄水量的保证，然而传统方式下蓄水量增大将提高水库建造成本和对蓄水安全性的要求。

采用分布式蓄水模式，在河流中游建立多个中间蓄水池，不仅可以减少主水库蓄水压力和成本，化整为零，还可以提高就近用水效率。

在大数据技术的支撑下，网络视频监控数据的存储模型可转向分布式的数据存储体系，提供高效、安全、廉价的存储方式。

（2）并行计算

并行计算是指采用多台计算机的计算资源，并行处理分布到各个节点的海量数据，提高数据处理的整体效率，这是目前提高大规模数据处理效率的有效手段。

并行计算主要分为 3 类，即 MPI、MapReduce 和 Dryad。

- MPI

MPI（Message Passing Interface，消息传递接口）是目前国际上并行计算领域最流行的 API 规范，由多家单位共同设计完成，易用性好、可移植性强、异步通信功能完备，是计算机集群、多处理器计算机、超级计算机进行高性能计算的常用技术。

在基于 MPI 的实现中，对于一个计算任务，一般需要划分为一组独立的计算部分，在初始化时对应生成一组进程，每一个进程完成一个计算部分，在不同节点上运行，进程之间通过集合通信或点对点通信方式进行数据交互，各个节点的计算结果最终汇总到主计算节点，完成同一个计算任务。

- MapReduce

MapReduce 是进行大规模数据处理的并行计算模型，由 Google 在 2004 年提出，应用于大规模集群。

Map（映射）和 Reduce（化简）是计算的两个阶段，前者通过调用 Map 函数实现一组键值到一组新键值的映射计算；后者采用 Reduce 函数对所有映射计算结果进行化简。

与 MPI 相比，MapReduce 在数据存储节点就地或就近完成 Map 或 Reduce 计算，减少了数据的网络传输压力。

- Dryad

Dryad 是微软在 2007 年提出的数据并行计算模型，与 MapReduce 相同，Dryad 也是通过在数据存储节点就地或就近完成相关计算的方式，减少数据的网络传输压力。

Dryad 采用 DAG（有向无环图）表示单个任务，按照 DAG 的方向依赖进行计算，计算类型相对于 MapReduce 更加丰富，计算结果可以通过 TCP Pipes、Shared-memory FIFOs 方式进行传输，避免冗余磁盘 IO 操作，传输手段更加高效。

第 2 章

海量视频模型

海量视频模型是海量视频处理与分析的基础，本章针对海量视频模型，重点介绍其基本理论和使用方法，包括 HSV 颜色模型、肤色模型、形状模型、人体可变形模型、混合高斯模型、概率图模型、感兴趣区域模型、视觉显著性模型、多分辨率模型、视觉词袋模型、视频语义模型等。

2.1 HSV 颜色模型

颜色模型是采用数学方法表示和处理视频图像信息的第一步，将颜色定义为特定空间的坐标值，不同的颜色空间定义可以得到不同的颜色模型，不同行业常用的颜色模型有 HSV、RGB、HSI、CHL、LAB、CMY 等，海量视频处理常用 HSV 颜色模型。

1. HSV 颜色模型的定义

如图 2.1 所示，在 HSV 模型中，每种颜色由色调（Hue，H）、饱和度（Saturation，S）和明暗度（Value，V）表示，对应于圆柱坐标系中的一个圆锥形子集。

色调 H 由绕 V 轴的旋转角给定，红色对应 0°，绿色对应 120°，蓝色对应 240°；每种颜色和它的补色相差 180°。

饱和度 S 取值从 0 到 1，等于颜色点到 V 轴的距离。

明暗度 V 取值从 0 到 1，对应于颜色点在 V 轴的投影位置。

圆锥顶面的半径为 1，顶面对应 V=1。顶面圆周上的颜色 V=1、S=1，这些颜色都是纯色。在圆锥的顶点处，即原点，V=0，为黑色。在圆锥的顶面中心处 V=1、S=0，H 无定义，为白色。从顶面中心点到原点，代表不同等级的灰色，S=0，H 无定义。

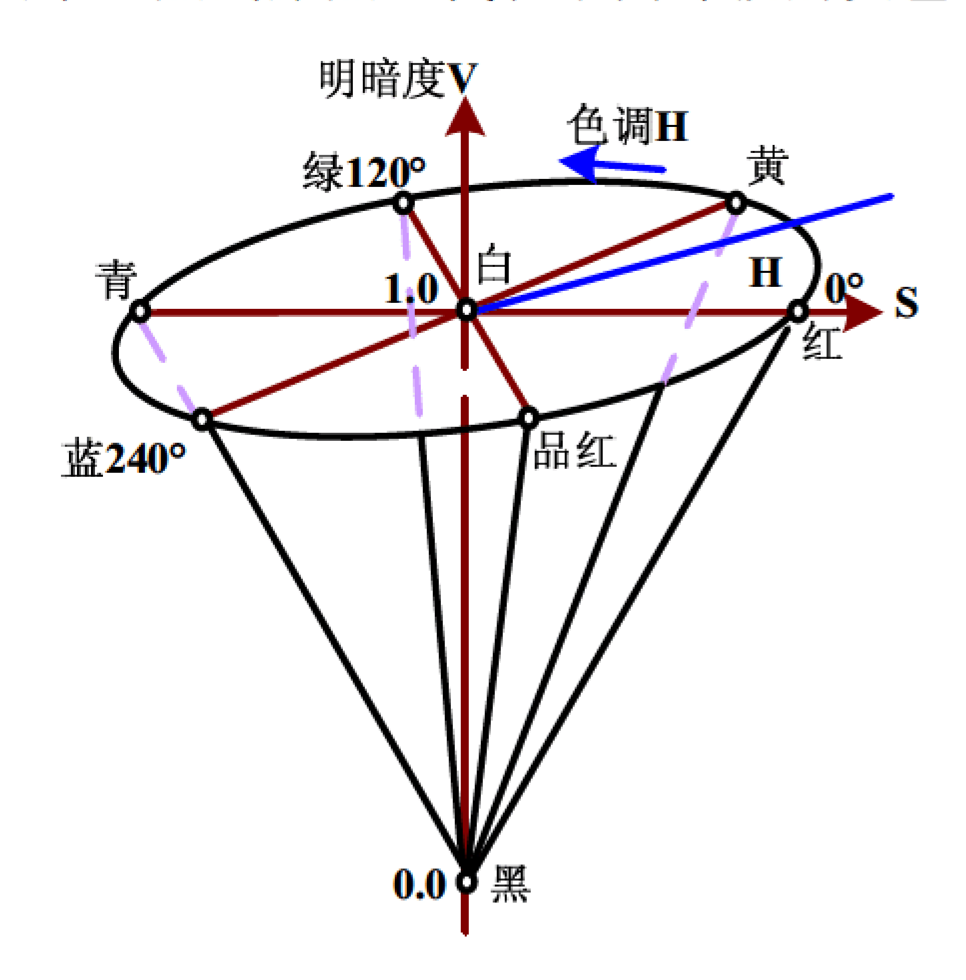


图 2.1　HSV 颜色空间

2. HSV 颜色模型的计算

设某颜色的 RGB 空间坐标为 (r,g,b)，对应 HSV 空间坐标为 (h,s,v)。其中，r、g、b、s、v 的取值范围为[0, 1]，h 的取值范围为[0, 360)。

❑ RGB 转换为 HSV

定义：

$$\max = \max(r,g,b)$$
$$\min = \min(r,g,b)$$

则有：

$$h = \begin{cases} 0^\circ & \max = \min \\ 60^\circ \times \dfrac{g-b}{\max-\min} + 0^\circ, & \max = r \text{ and } g \ge b \\ 60^\circ \times \dfrac{g-b}{\max-\min} + 360^\circ, & \max = r \text{ and } g < b \\ 60^\circ \times \dfrac{b-r}{\max-\min} + 120^\circ, & \max = g \\ 60^\circ \times \dfrac{r-g}{\max-\min} + 240^\circ, & \max = b \end{cases}$$

$$s=\begin{cases}0, & \max=0\\ \dfrac{\max-\min}{\max}=1-\dfrac{\min}{\max}, & \text{其他}\end{cases}$$

$$v=\max$$

在 MATLAB 中，有对应的转换函数 $HSV=rgb2hsv(RGB)$。

- HSV 转换为 RGB

$$h_i=\left\lfloor\frac{h}{60}\right\rfloor \bmod 6$$

$$f=\frac{h}{60}-h_i$$

$$p=v\times(1-s)$$

$$q=v\times(1-f\times s)$$

$$t=v\times(1-(1-f)\times s)$$

$$(r,g,b)=\begin{cases}(v,t,p), & h_i=0\\ (q,v,p), & h_i=1\\ (p,v,t), & h_i=2\\ (p,q,v), & h_i=3\\ (t,p,v), & h_i=4\\ (v,p,q), & h_i=5\end{cases}$$

在 MATLAB 中，有对应的转换函数 $RGB=hsv2rgb(HSV)$。

3. HSV 颜色模型的应用

如图 2.2 所示，色调信号 H 从 0° 到 360° 变化，其中 S=1、V=1，对应颜色依次从红色、黄色、绿色到蓝色逐渐变化。



图 2.2　色调信号变化效果图

人的肤色主要集中在黄色和红色区域，采用色调信号，HSV 模型可以用于人脸检测，

实现人脸分割，如图 2.3 所示。

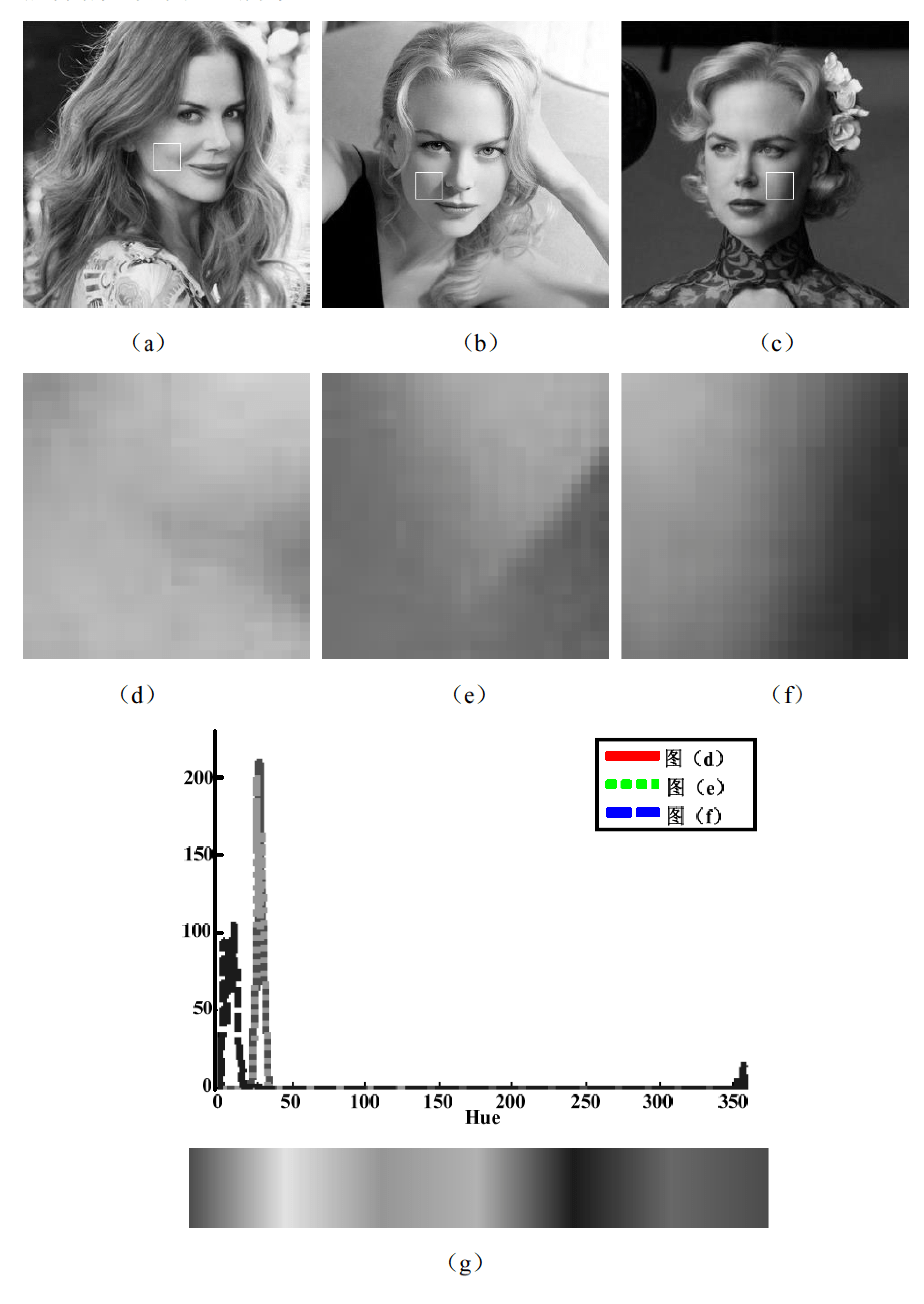


图 2.3　基于 HSV 模型的人脸检测

图 2.3（a）、（b）、（c）对应在不同角度和光照条件下的人脸图像，各图中的白框表示提取分析的色块，分别显示为图 2.3（d）、（e）、（f）。图 2.3（g）是各个色块对应 H 分量的直方图。在不同条件下，人脸区域的色块的 H 分量直方图分布具有显著特性：在一定范围内集中分布。利用该特性，可以设计算法，高效实现人脸的检测与分割。

2.2　肤色模型

将皮肤颜色映射到 YCbCr 空间，在 CbCr 二维平面中肤色近似成一个椭圆分布。

1. YCbCr 空间的定义

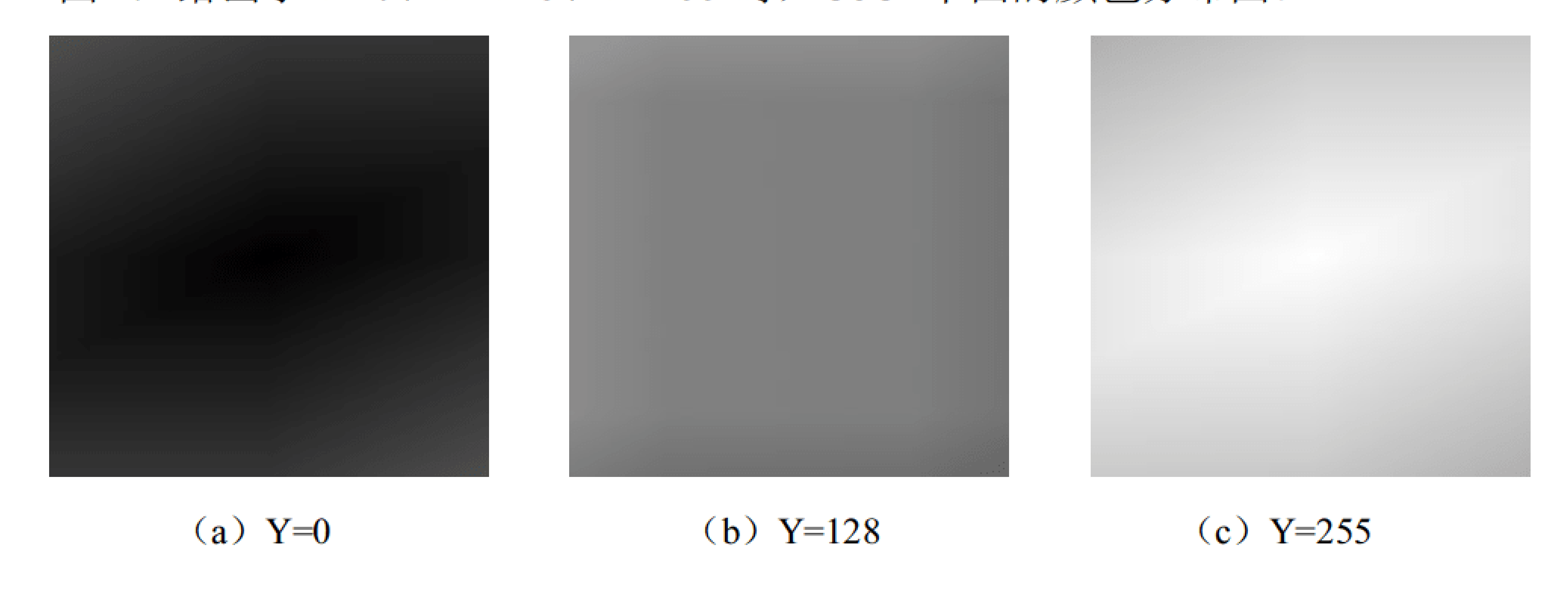
在 YCbCr 颜色空间中，Y 代表亮度，为 RGB 信号的加权平均值。色度采用 Cb 和 Cr 表示，Cb 反映 RGB 信号中蓝色部分与亮度值之间的差异，Cr 反映 RGB 信号中红色部分与亮度值之间的差异。

YCbCr 的具体实现有多种形式，可以根据具体情况优化选择。下式可以实现 YCbCr 与 RGB 空间的转换，当 RGB 各分量在[0, 255]时，转换的 Y 属于[0.0, 255.0]，Cb、Cr 属于[-128.0, 127.0]。

$$[Y\ Cb\ Cr]=[R\ G\ B]\begin{bmatrix} 0.299 & -0.168935 & 0.499813 \\ 0.587 & -0.331665 & -0.418531 \\ 0.114 & 0.50059 & -0.081282 \end{bmatrix}$$

2. YCbCr 空间的颜色分布

图 2.4 给出了 Y=0、Y=128、Y=255 时，CbCr 平面的颜色分布图。



（a）Y=0　　（b）Y=128　　（c）Y=255

图 2.4　CbCr 平面的颜色分布图

对于不同的 Y 值，CbCr 平面的颜色分布具有相对固定的特征。各图左上角对应（Cb

= -128, Cr = -128），表现为不同程度的绿色；各图右上角对应（Cb = 127, Cr = -128），表现为不同程度的蓝色；各图左下角对应（Cb = -128, Cr = 127），随着亮度的变化从红色、橙色变化到黄色。

3. 肤色模型的应用

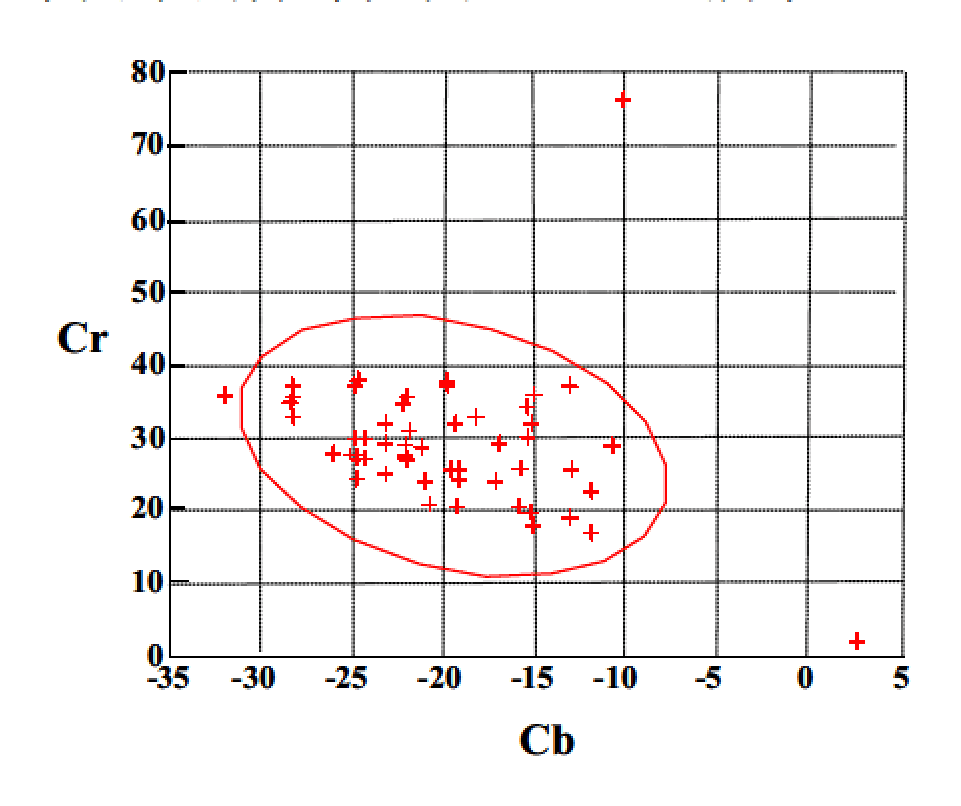
肤色模型的应用包括两个步骤，即肤色模型的构建和使用。

❑ 肤色模型的构建

设有如图 2.5（a）所示的样本图像，包含人脸及手臂等肤色像素。为了构建肤色模型，从图像中采样皮肤像素点，并将其从 RGB 空间投影到 YCbCr 空间。各采样点的位置在图 2.5（a）中用红色的交叉表示，在 CbCr 平面中的分布如图 2.5（b）所示。




（a）样本图像



（b）CbCr 平面内的样本分布

图 2.5　肤色模型的构建

在 CbCr 平面内，肤色样本点的分布近似于高斯椭圆。利用样本点信息，可以求取包含大部分样本点的椭圆曲线，可以用椭圆参数模型描述。

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} Cb - cb_0 \\ Cr - cr_0 \end{bmatrix}$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

其中，θ是长轴对应的旋转角度；a、b 分别对应长、短轴半径；（cb_0, cr_0）对应样本的平均值。

❑ 肤色模型的使用

根据肤色模型，可以得到如下的判断准则：

$$D(Cb,Cr)=\begin{cases}1, & \frac{x^2}{a^2}+\frac{y^2}{b^2}\le 1\\ 0, & \text{其他}\end{cases}$$

对任一像素点，如果其在 CbCr 平面内的投影在椭圆曲线内，则判断为皮肤点；反之，则不是。根据这个判断准则，可以得到图 2.6 所示的肤色判断结果图。



图 2.6　肤色判断结果

通过建立肤色模型，可以判断像素点是否属于皮肤。从图 2.6 所示的肤色判断结果可以看出：

- 肤色判断基本准确，可以将人脸和手臂部分有效地分割出来。
- 存在一定误差，女士的头发由于与肤色相似，被错分为皮肤。
- 对光照有一定鲁棒性，可有效识别脸上的阴影区域。
- 对光照的鲁棒性有限，左一人物的左眼亮度过暗，无法有效识别。

在实际使用中，该肤色模型可以作为一个预处理环节，与其他信息融合，实现高效、精确的皮肤检测与分割。

2.3　形状模型

基于主动形状模型（Active Shape Models，ASM）的目标检测方法广泛应用于海量视频处理中，利用训练所建模型与新数据的匹配，实现目标的检测与定位。

1. 主动形状模型的建模

为了建立目标形状的统计模型，需要一些典型样本图像。在包含目标的样本图像中，人工标记目标的形状信息。将标记的数据作为训练的样本，根据其统计特性建立模型。ASM 生成过程如下。

❑ 选择合适的标记点

标记点是 ASM 的基础，合适的标记点便于检测和定位。如图 2.7 所示，一般选择目标边缘的角点、交叉点等特征点。为了避免标记点过于稀疏，往往在这些点之间沿着目标边界，等距地选取中间插值点作为辅助。

特征点、插值点及其邻接关系共同表征目标的形状。

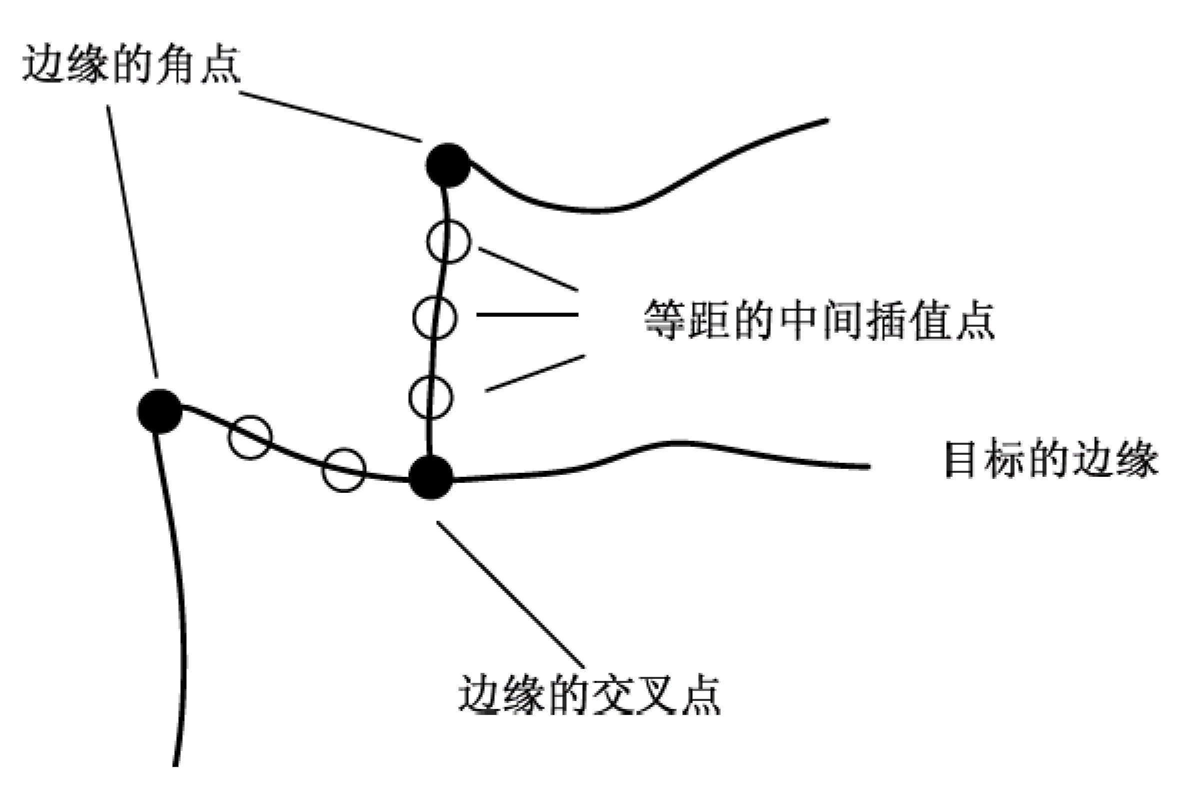


图 2.7　合适的标记点

❑ 生成形状特征向量

记录选取的标记点及其连接顺序，得到有序点列：

$$\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$$

将这些点的坐标分组，依次串联起来，得到表征目标形状的特征向量：

$$X=(x_1\ldots,x_n,y_1,\ldots,y_n)^T$$

为了使该模型与其在图像中的位置、角度和尺度无关，需要进行归一化操作，将不同样本的图像坐标变换到统一的坐标系。

❑ 建立形状的统计模型

每个训练样本对应 $2n$ 维特征空间中的一个点，而目标在这个 $2n$ 维特征空间内的分布就是其形状模型，可以根据训练样本估计其特征分布。

为了简化分析，使用主成分分析（PCA）方法将特征空间降维。每个训练样本 X，通过 PCA 降维，可以近似为：

$$X\approx\overline{X}+Pb$$

其中，$\overline{X}$ 对应平均模型；$P=(p_1,p_2,\ldots,p_t)$ 包含样本协方差矩阵中特征值最大的 t 个特征向量，b 是一个 t 维向量：

$$b = P^T\left(X - \overline{X}\right)$$

b 可以看作是形状模型的形变参数，通过改变 b，可以得到形状模型不同的变形实例。假设第 i 维参数 b_i 在样本集下对应的特征值为 λ_i，将 b_i 的变化范围限制在 $\pm 3\sqrt{\lambda_i}$，可以将模型的形变控制在与训练样本相似的范围内。

2. 主动形状模型的匹配

已知目标 ASM 模型及测试点列，需要利用匹配算法求取测试点列对应的形变参数，并依此识别和定位目标。由于 ASM 模型建立在归一化的坐标系下，图像坐标系下的测试点列，一般需要经过平移、旋转和缩放等坐标变换，才能和目标模型进行匹配。

❑ 坐标变换

模型的坐标变换为：

$$X = T_{X_t,Y_t,s,\theta}\left(\overline{X} + Pb\right)$$

函数 $T_{X_t,Y_t,s,\theta}$ 实现平移、旋转和缩放：

$$T_{X_t,Y_t,s,\theta}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} X_t \\ Y_t \end{pmatrix} + \begin{pmatrix} s\cos\theta & -s\sin\theta \\ s\sin\theta & s\cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

其中，(X_t,Y_t,s,θ) 为坐标变换参数，(X_t,Y_t) 对应平移，s 对应缩放，θ 对应旋转。

❑ 匹配过程

已知 ASM 模型，对于图像中的一个测试点列 Y，求取其最优的形变参数 b 和对应的坐标变换参数 (X_t,Y_t,s,θ)。该匹配过程可以表述为下式的最小化问题：

$$\left|Y - T_{X_t,Y_t,s,\theta}\left(\overline{X} + Pb\right)\right|^2$$

可以使用算法 2.1 所示的迭代方法求解。

算法 2.1　ASM 模型匹配算法

过程：1. 初始化形变参数 $\boldsymbol{b}$=**0**，对应平均模型；
2. 利用模型及其参数 $\boldsymbol{b}$，生成模型实例 $X = \overline{X} + Pb$；
3. 求取模型实例 X 与测试点列 Y 之间的最佳坐标变换参数 (X_t,Y_t,s,θ)；
4. 利用最佳坐标变换参数，将 Y 映射到归一化坐标下，得 $\hat{Y}$：

$\hat{Y} = T^{-1}_{X_t,Y_t,S,\theta}(Y)$；

5. 求取 $\hat{y}$ 对应的形变参数 $\boldsymbol{b}$：　$b = P^T(\hat{Y} - \overline{X})$；
6. 判断收敛性，如果没有收敛，跳转到步骤 2；如果收敛，算法结束。此处收敛的含义为当次迭代没有使形变参数或者坐标变换参数产生显著变化。

3. 主动形状模型的应用

❑ 建模

利用如图 2.8 所示的手掌图像，建立手掌边界 ASM 模型。

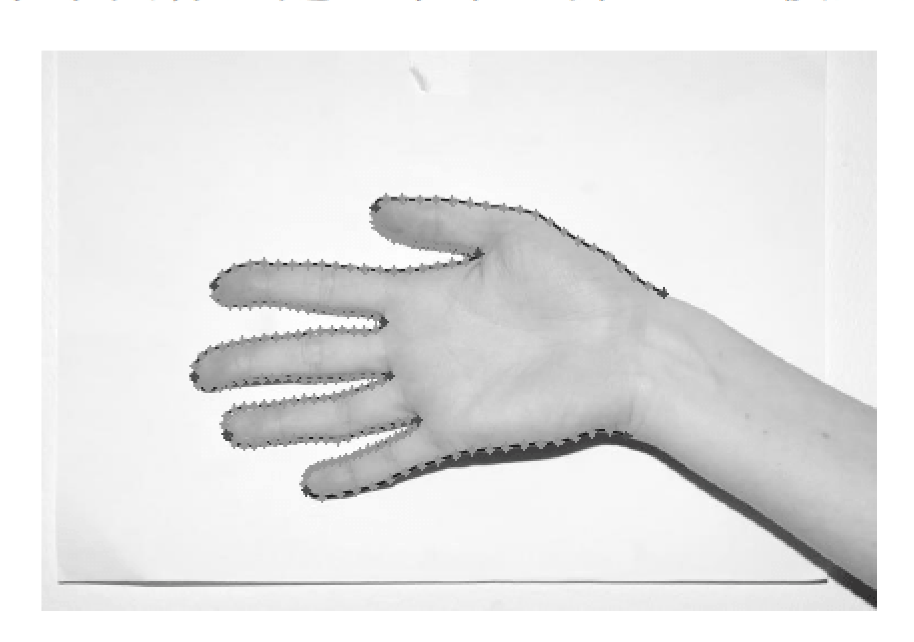
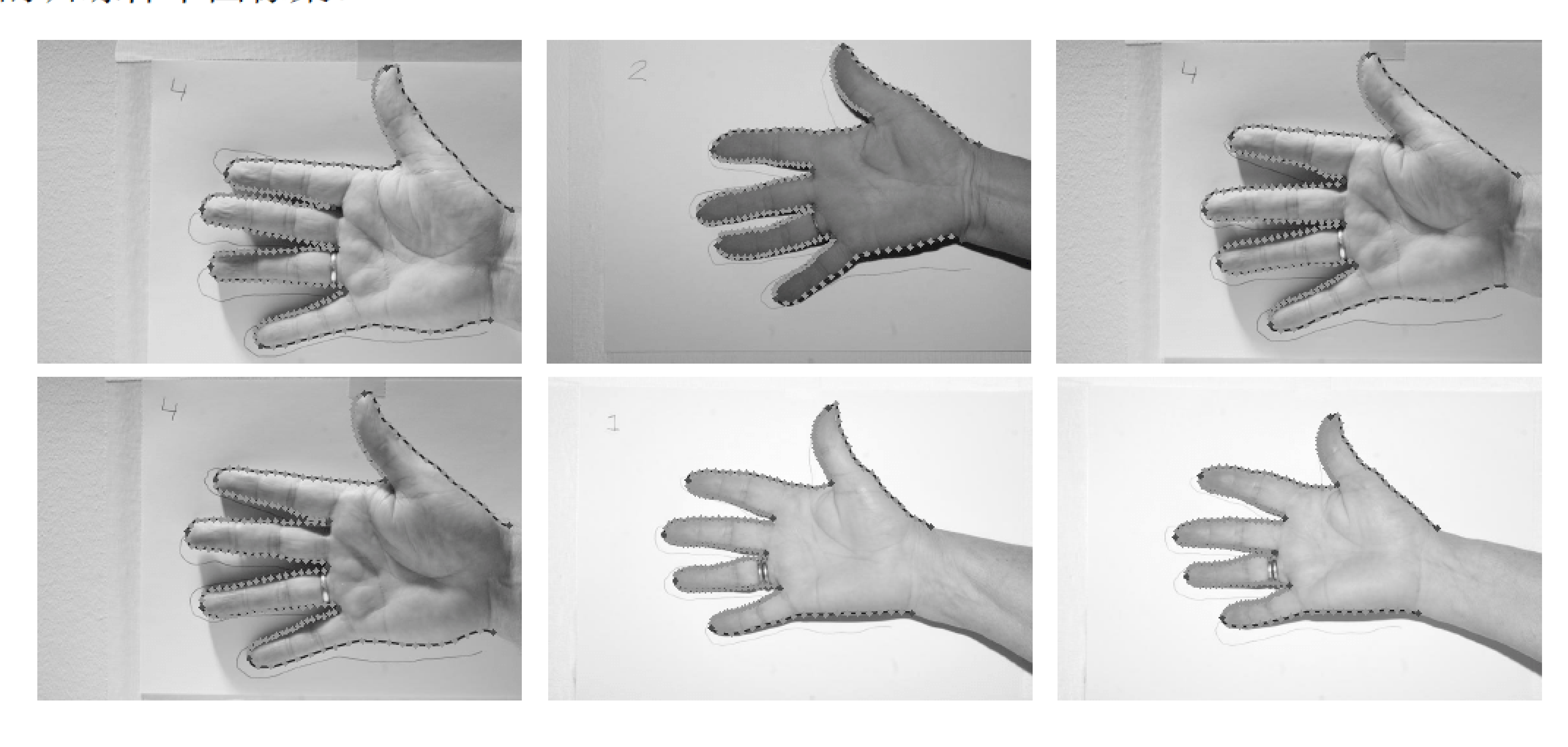


图 2.8　手掌边界建模

图 2.8 中红色和绿色的点均为模型的标记点，红色的点是易于检测的特征点，绿色的点是特征点之间的插值点。这些点及其邻接关系共同表征手掌形状，图 2.9 是标注好的训练样本图像集。



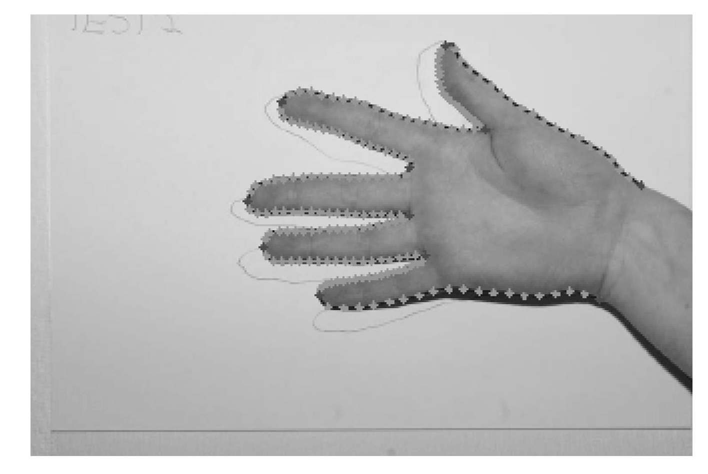
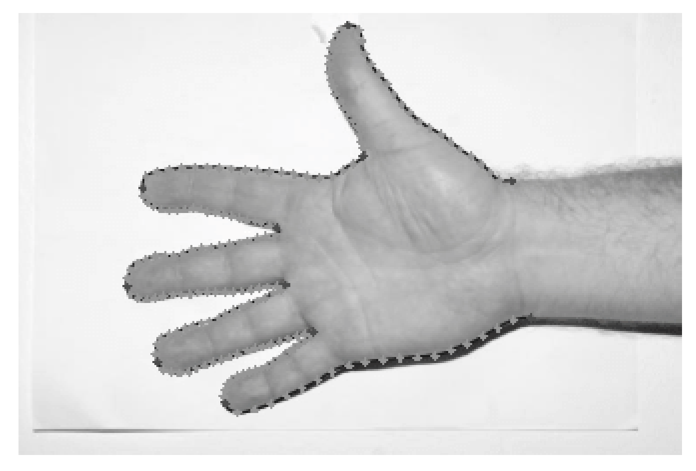
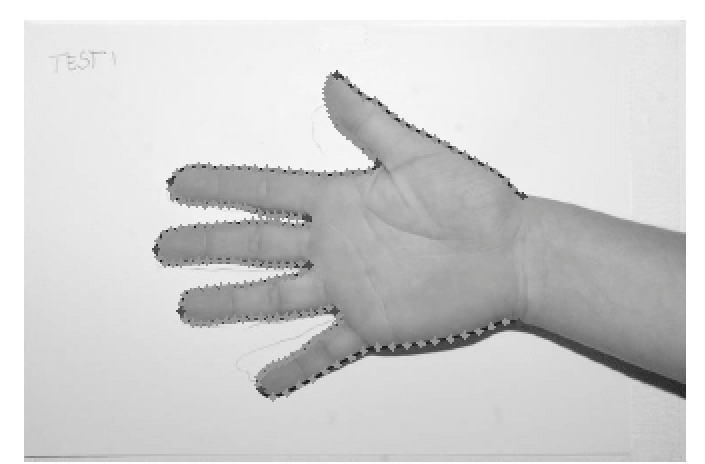
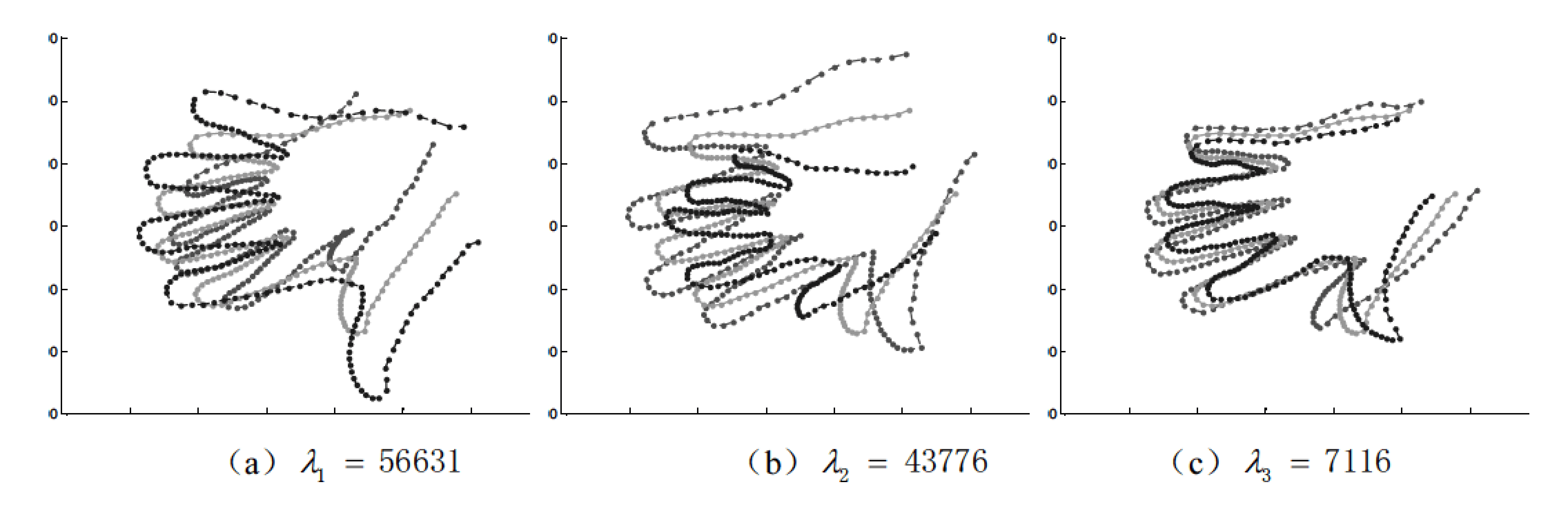
  

图 2.9　手掌建模训练样本

利用如图 2.9 所示的训练样本，可以得到手掌模型，如图 2.10 所示。其中绿色的点列表示模型的平均形状 $\bar{X}$，红色和蓝色的点列表示某个 b_i 取值为 $\pm 3\sqrt{\lambda_i}$ 时的形状。



（a）$\lambda_1 = 56631$　　（b）$\lambda_2 = 43776$　　（c）$\lambda_3 = 7116$

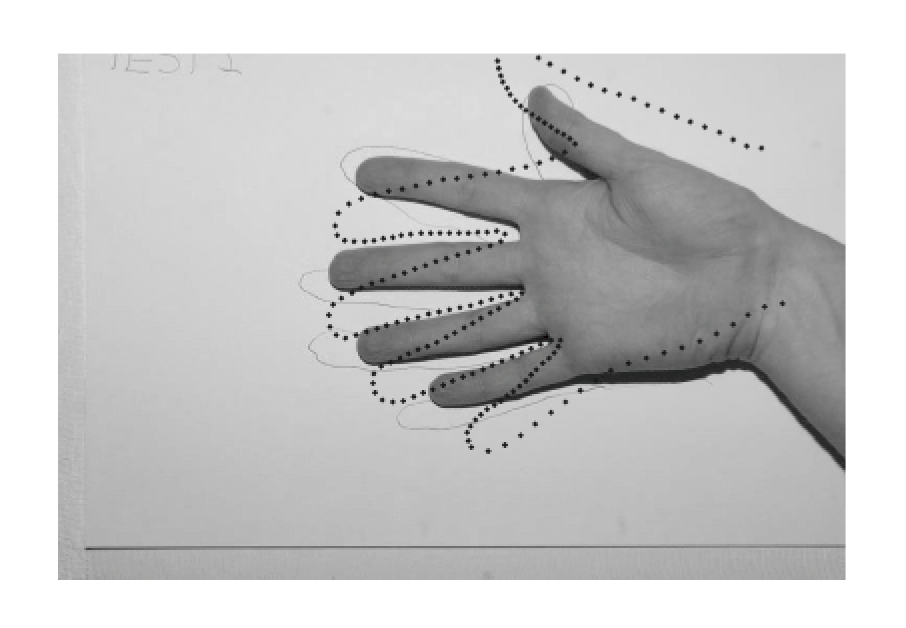
图 2.10　手掌模型及其变形

可以看到，前两个特征值比较大，形变比较剧烈，第三个特征值较小，形变也较小。

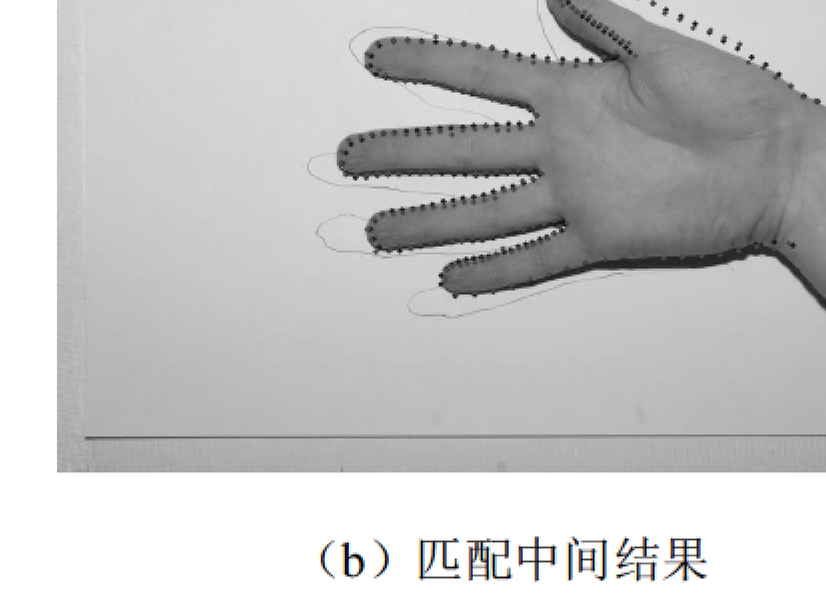
❑ 匹配

对某测试图像，选定一个初始位置，利用前述的迭代算法，可以求取不断优化的形变参数，实现目标的检测与定位。

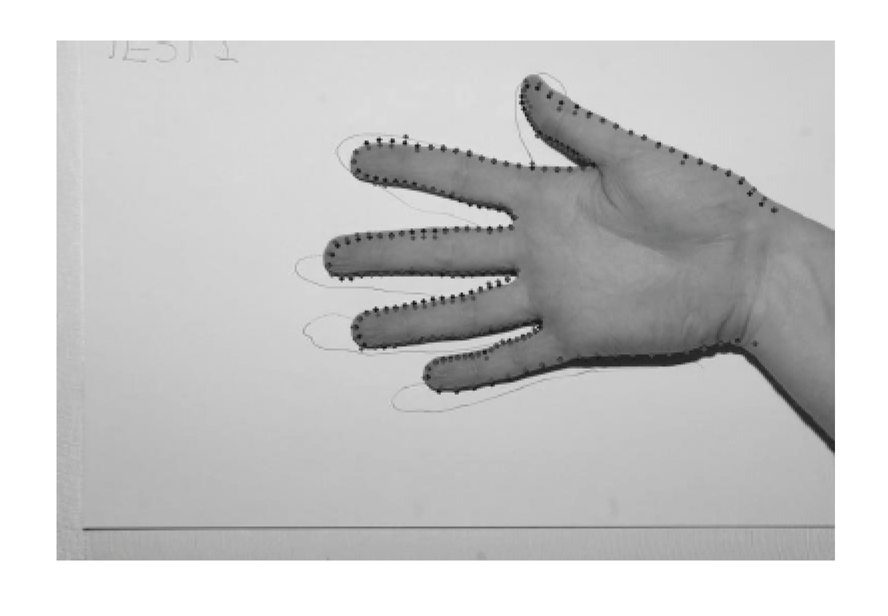
在图 2.11（a）中，蓝色点列代表迭代初始位置，由人工给定。在图 2.11（b）、（c）中，蓝色点表示某步迭代前的标记点位置，红色点为迭代后的标记点位置，当迭代不能进一步优化时，输出检测结果。图 2.11（c）为最终匹配结果，图 2.11（d）为分割效果，最终结果能够对手掌区域进行合理的检测及定位。



（a）匹配初始位置



（b）匹配中间结果



（c）匹配最终结果



（d）检测分割效果

图 2.11　匹配过程及结果

4. 讨论

形状模型有其适用的条件和局限性，在其适用的范围内，该模型有显著优势。在适用的范围外，其效果可能较差，甚至不能使用。

形状模型适用的条件有：

- ❑ 目标具有显著的形状特征，如手掌、人脸。
- ❑ 可以找到一定数量的典型样本，实现人工标记。
- ❑ 可以基本准确地确定目标初始位置，否则迭代算法容易陷入极小局部，难以收敛。
- ❑ 目标的拓扑结构不能发生变化，目标上必须有明显的标记点。当目标具有多种形态时，该方法一般不适用，如树、烟、水等。

形状模型只能在训练样本限定的范围内变化，提高样本多样性，可以有效提升模型的适用范围。对于视频跟踪，该算法具有一定优势，前一帧的检测结果可以作为下一帧的初始位置。

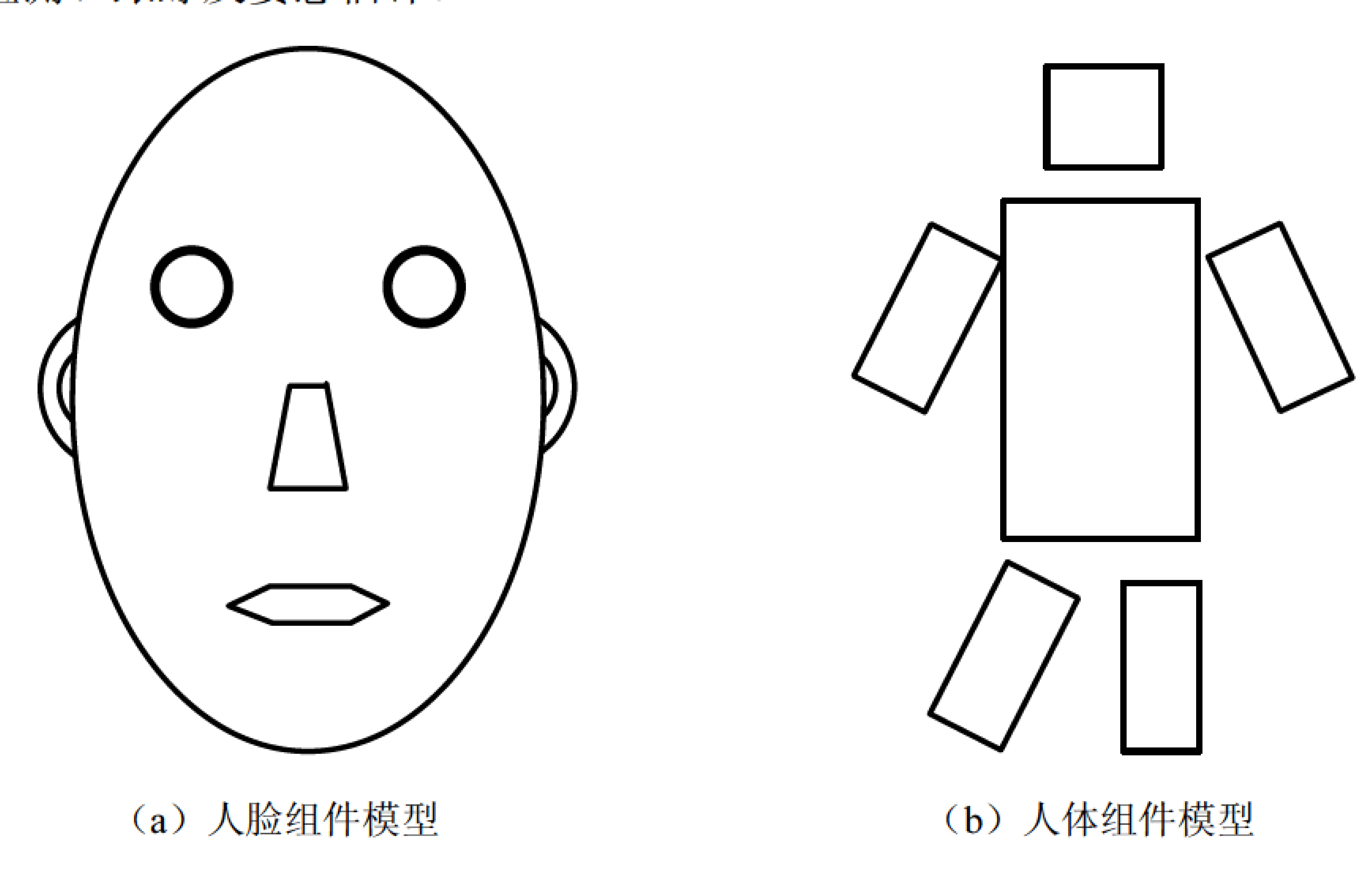
2.4 人体可变形模型

运动目标的检测与跟踪是海量视频处理的重要课题，本节介绍人体可变形模型在人的检测、跟踪及姿态估计中的应用。

1. 人体可变形模型的背景

人体可以表示为不同组件及其相对关系的综合，人的运动体现为不同组件相对关系的变化。

图 2.12 给出了人脸和人体的组件模型。人脸由眼睛、鼻子、嘴巴、耳朵等组件依照一定相对关系组成。人体可以分解为躯干、四肢和头部。人体在运动过程中，其组件的相对位置变化遵循一定的规律。人体可变形模型，就是利用这些固有的约束，高效地实现人的检测、跟踪及姿态估计。



（a）人脸组件模型　　（b）人体组件模型

图 2.12　人脸和人体的组件模型

2. 人体可变形模型的建模

❑ 组件的建模

组件的识别与定位是人体可变形模型的基础。组件识别有多种方法，可以是利用颜色模型的皮肤检测；可以是利用 SIFT 算子的关键点检测；也可以是基于 HOG 算子的区域检测。

HOG 算子在组件建模中应用较多，其检测的评价指标可以表示为模型与局部特征的卷积：$w_i \cdot \phi(I, l_i)$。其中 w_i 是组件的模型，$\phi(I, l_i)$ 是目标图像局部区域的 HOG 特征。

❑ 结构的建模

结构就是组件间的相对位置关系。组件在目标中的相对位置参数，可以通过对样本的统计得出，往往归结为最大似然估计问题。

❑ 模型的推理

模型的推理可以理解为一个优化问题，求取最优的组件位置，综合局部检测算子的评价和各个组件之间的相对位置信息，使得总体的评价最高。模型的推理，往往归结为最大后验估计问题。

3. 人体可变形模型的应用

人体可变形模型主要应用于人的检测、跟踪、姿态估计。

使用人体可变形模型可以实现人的检测。由于使用组件及其相对位置关系，检测精度和稳定性较高。由于需要对组件进行建模，于是需要人体在图像中占有一定的像素数目。当图像中的人很小而无法有效检测各个部分时，该方法的效果并不理想。

对于人的跟踪，该模型有两种方法。第一种，将前一帧的结果作为当前帧的初始值，在此基础上求出优化的结果；第二种，各帧独立检测，通过后处理实现跟踪，如对目标运动的错误检测、低通滤波。

由于人体可变形模型已经求出各个组件及其相对位置，于是可以估计人体姿态信息，如站立、行走、弯腰等。

2.5　混合高斯模型

混合高斯模型是视频图像处理的基础模型，本节介绍混合高斯模型的定义、参数求取和应用实例。

1. 混合高斯模型的定义

混合高斯模型是高斯模型的扩展，是多个高斯模型的线性组合。

假设样本 x 是 z 的实例，z 可能属于 K 个类别中的任何一类，并且属于第 k 类的概率为 π_k；每个类别的样本均满足一个高斯分布 $N\left(x \middle| \mu_k, \Sigma_k\right)$，于是 x 的概率分布表示为：

$$p(x) = \sum_z p(z) p(x|z) = \sum_{k=1}^{K} \pi_k N\left(x \middle| \mu_k, \Sigma_k\right)$$

图 2.13 给出了 K 为 3 时二维空间中某混合高斯分布的示意图，单一的高斯模型无法精确地刻画该分布，混合高斯模型能够提高模型的适应性。

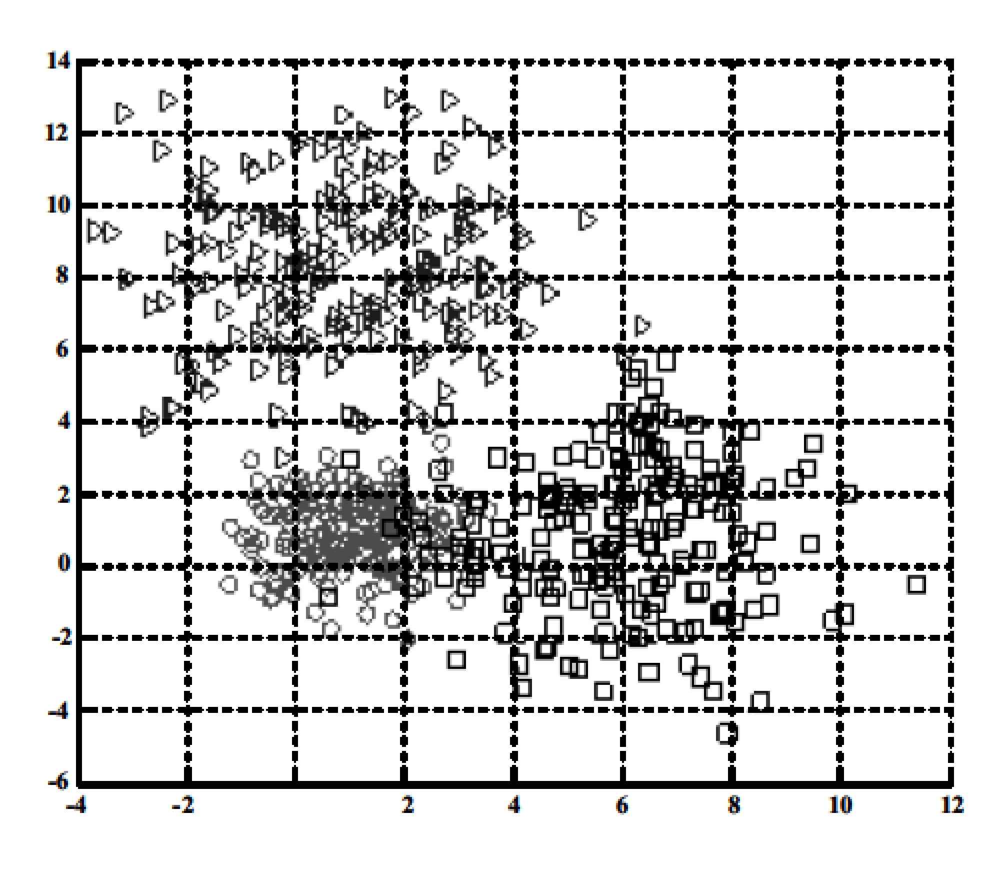


图 2.13 混合高斯模型示意图

2. 混合高斯模型的参数求取

假设有 N 个样本点{x1,⋯,xN}，以此建立混合高斯模型，即求取高斯模型的各个参数，包括π_k、μ_k、Σ_k。

混合高斯模型的求解常用 EM（Expectation Maximization）估计法，通过逐步迭代策略实现模型参数的估计和优化，包括两个迭代步骤，即 E 步骤和 M 步骤。EM 估计法的具体实现流程如下。

算法 2.2 EM 算法

过程：1. 初始化

初始化π_k、μ_k、Σ_k，可以随机选取，也可先用 K-Means 聚类估计。

2. E 步骤

计算每个样本属于 K 个类别的概率

$$\gamma(z_{nk})=\frac{\pi_k N(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j,\Sigma_j)}$$

3. M 步骤

计算各类别的模型参数

$$\mu_k^{new}=\frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})x_n$$

$$\Sigma_k^{new} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}$$

$$N_k = \sum_{n=1}^{N}\gamma(z_{nk})$$

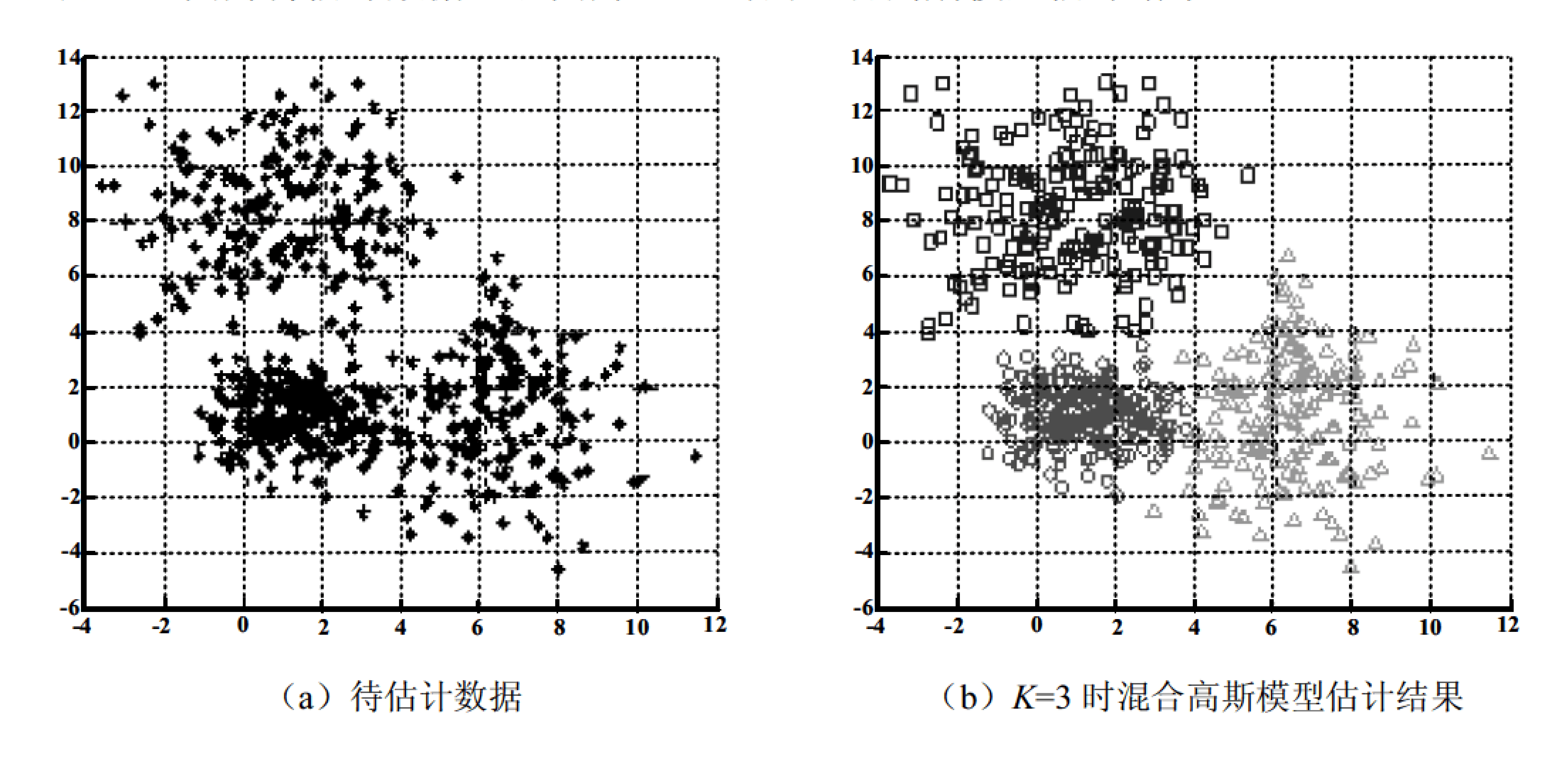
4. 收敛性判断

计算 X 属于该模型的评价函数

$$\ln p(X|\mu,\Sigma,\pi) = \sum_{n=1}^{N}\ln\left\{\sum_{k=1}^{K}\pi_k N(x_n|\mu_k,\Sigma_k)\right\}$$

当该函数趋于稳定，或模型参数不再变化时，算法收敛，迭代结束；否则，转到步骤 2。

如图 2.14 所示，利用 EM 估计法可以求取混合高斯模型的参数，该方法假设 K 值已知，左图为待估计数据，右图为 K=3 时的混合高斯模型估计结果。



（a）待估计数据　　（b）K=3 时混合高斯模型估计结果

图 2.14　基于 EM 方法的高斯模型估计

3. 混合高斯模型的应用实例

混合高斯模型最典型的应用是监视场景的背景建模与运动目标检测，通过建立背景模型，检测运动目标。

如图 2.15 所示，在某监控场景中，摄像机位置固定，对每个像素建立独立的混合高

斯模型，根据实时的视频图像数据动态更新模型参数。对于新来的像素，通过判断其属于背景的概率来分辨运动物体和静态场景。

如图 2.15 所示，左图为当前帧的图像；右图为基于混合高斯模型的运动目标检测效果，白色表示其为运动目标的概率大，黑色表示其为背景的概率大。

该模型可以较好地估计运动目标，可以适应光照的缓慢变化。在运动目标（如汽车）尾部存在一定的拖尾现象；对于运动目标上大面积相同颜色的区域也存在漏检现象，这些问题可以通过后处理环节来解决。



（a）视频图像　　（b）运动目标检测效果

图 2.15　基于混合高斯模型的运动目标检测

2.6　概率图模型

概率图模型是概率论和图论的综合，采用图的方式来探究随机变量之间的条件独立性，给出随机变量的联合概率分布。

1. 概率图模型的表述

概率图模型由节点和连接组成，采用节点表示随机变量，采用节点间的连接表示随机变量之间的条件独立性。概率图的整体表述随机变量的联合概率分布。概率图模型可以分为有向图模型、无向图模型和因子图模型。

❑ 有向图模型

有向图模型即贝叶斯网络，其图结构是有向无环图，有向的连接表示因果关系，节点上存储的往往是随机变量之间的条件概率表格，其联合概率分布可依全概率公式给出。图 2.16（a）是某有向图模型的实例。

❑ 无向图模型

无向图模型即马尔科夫网络，其图结构是无向图，无向的连接表示变量之间的相关性，其联合概率分布定义为子集势能函数的乘积。图 2.16（b）是某无向图模型的实例。

❑ 因子图模型

因子图模型即双相图模型，包含两组不同的节点。一组节点表示随机变量，一组节点表示因子，连接只存在于两种不同的节点间，其联合概率分布采用因子乘积表示。图 2.16（c）是某因子图模型的实例。

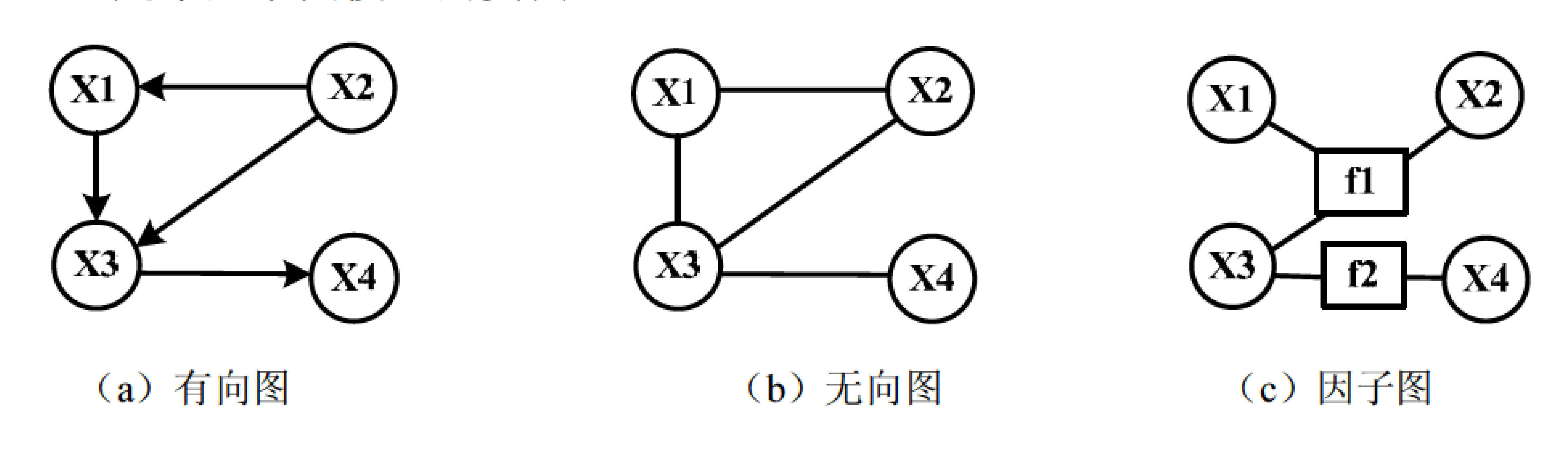


图 2.16　概率图模型的表述

2. 概率图模型的学习

为了使用概率图模型，必须给定模型中的参数。概率图模型的参数可以人工给出，如由贝叶斯网络构成的专家系统给出，其参数往往是专家经验的汇总，其中的连接表示因果关系，条件概率表征可能性的大小。当概率图的规模变得很大时，人工指定参数的方法难以实现，于是就有基于学习的自动参数生成法。概率图模型的学习，就是基于样本集合推断概率图中的模型参数。

❑ 产生式学习（generative learning）

给出目标模型的样本集合，求取概率图模型近似目标模型的联合概率分布。由于求取的是产生样本的模型，于是叫做产生式学习。求出概率图模型之后，其他感兴趣的量就可以方便得出。在实际应用中，尤其是在样本数目有限的条件下，该方法很难得到满意的结果。

❑ 分辨式学习（discriminative learning）

当使用概率图模型的目的不是完整的建模，而是分类，即推断某些变量的类别属性时，可以使用分辨式学习。该方法直接优化分类误差，无须对样本模型整体建模，便可以得到更好的分类效果。

3. 概率图模型的推理

推理就是在可以部分观察到某些变量的基础上，利用概率图模型，计算某些变量的分布，或者某些变量的最大似然估计。

- 准确推理

如果图的结构是树，那么 BP 算法（Belief Propagation）可以准确求解其边缘分布。对于任意的图结构，可以使用交叉树算法（Junction Tree Algorithm）求解。在图结构复杂、规模庞大时，准确求解变得非常慢，通过近似求解算法可以实现近似的、高效的推理计算。

- 近似推理

近似推理法主要有变分算法（Variational Methods）、循环信息传递法（Loopy Message Passing）、基于采样的方法（Sampling Methods）。

2.7　感兴趣区域模型（ROI）

感兴趣区域模型（Region of Interest，ROI）在视频图像处理中有着广泛的应用，本节介绍感兴趣区域模型的定义和应用。

1. 感兴趣区域模型（ROI）的定义

感兴趣区域（ROI）是通过预处理（人工或者自动）选出的特定区域，通常对 ROI 数据进行特别处理，提取 ROI 对于提高数据利用率具有重要意义。

感兴趣区域的应用价值如下：

- 节省存储空间，对于感兴趣区域之外的区域，可以采用压缩比更高的算法，实现对数据的高效存储。
- 节省处理时间，仅对感兴趣区域进行重点处理，可以有效节省算法运行时间。
- 提高处理效果，在节省空间、节省时间的条件下，可以综合多种算法，提高处理效果。

如图 2.17 所示，对于假想的金库入口监控图像，金库入口区域是处理重点，可以选择方框所示区域为感兴趣区域。重点区域占整个图像的面积不到一半，对该区域的针对性处理，可以有效提高存储和计算的效率。

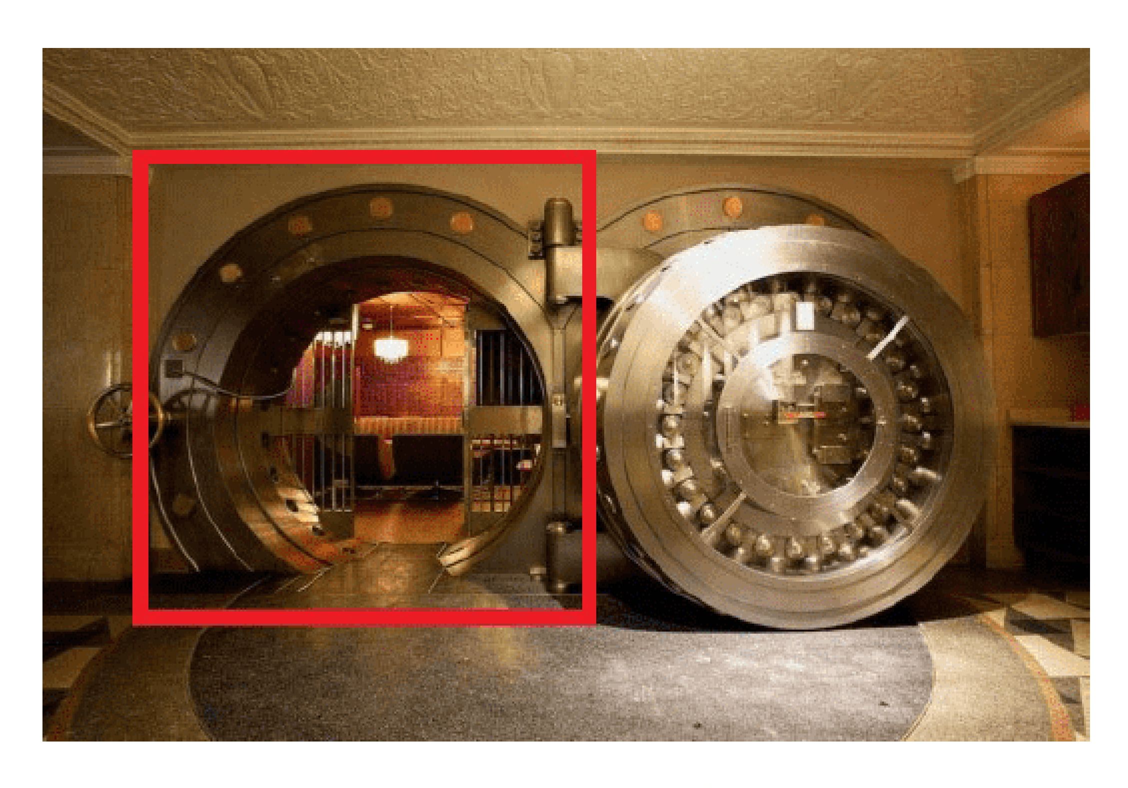


图 2.17　ROI 示意图

2. 感兴趣区域模型（ROI）的应用

ROI 模型在空域和时域的应用示例中，ROI 区域可以通过人机交互选定，也可以通过一定的预处理算法自动提取。

- 空域 ROI 的应用示例

如图 2.18 所示，在道路监控视频图像中，包含天空、高架桥、树木、绿化带和道路区域。如果需要设计监控程序监控道路车辆运行情况，那么可以选择图中标示的感兴趣区域，ROI 之外存在感兴趣车辆的概率几乎为 0。这个 ROI 区域，可以人工选定，也可以通过分析运动目标得出。



图 2.18　空域 ROI 示意图

- 时域 ROI 的应用示例

如图 2.19 所示，假设某校园内的监控摄像头，通过背景建模，可以计算得到视频图像的场景变化指数，即颜色/灰度变化超过一定阈值的像素数目。考虑到光照和树叶摇晃

等干扰因素的影响，选择一定阈值，如图 2.19 中虚线所示，虚线以下区域认为场景变化较小，予以简化处理；虚线以上区域确定为 ROI，予以重点分析，这样可以缩减约 2/3 的处理时间和存储空间。

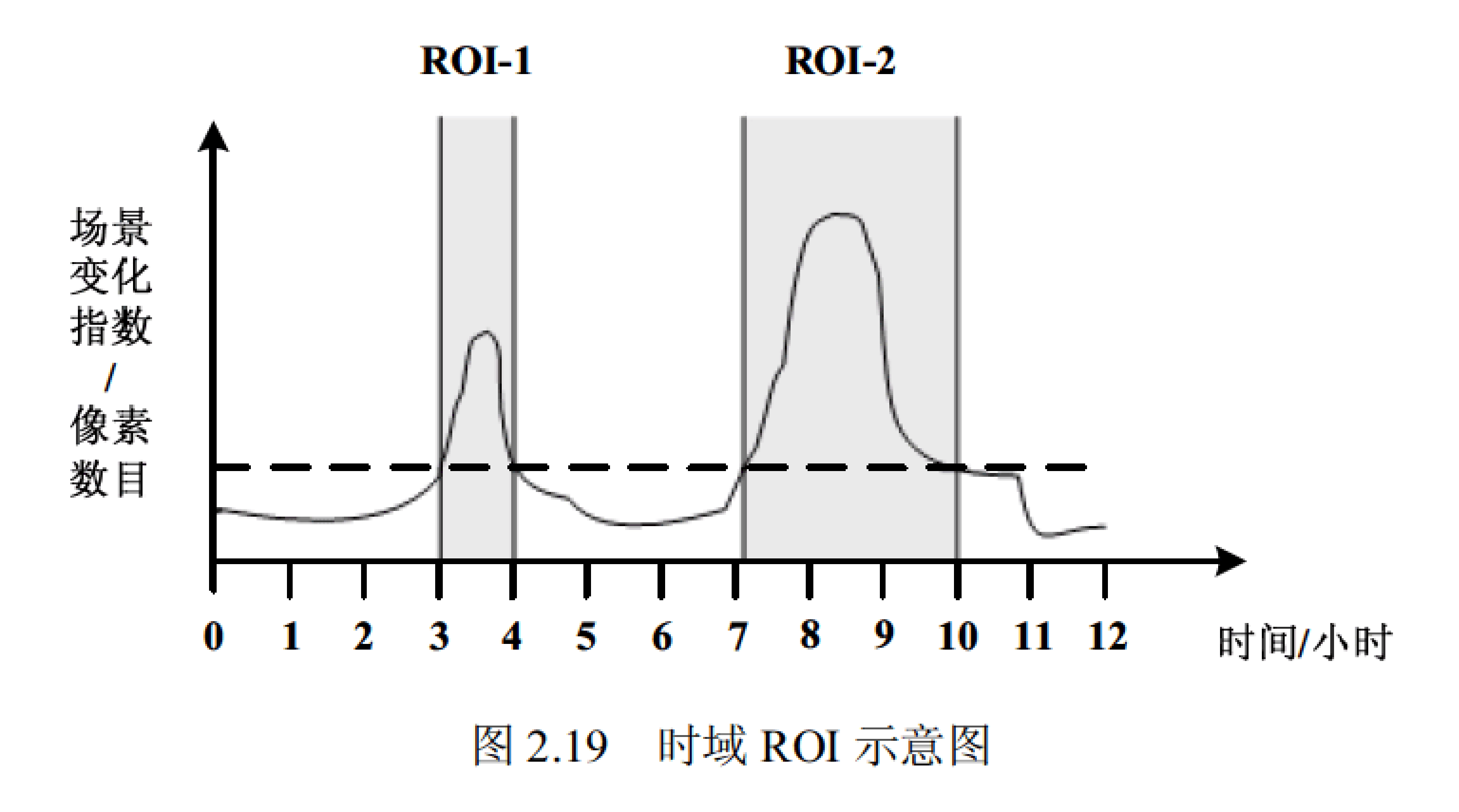


图 2.19 时域 ROI 示意图

2.8 视觉显著性模型

人类的视觉系统可以通过选择性注意机制有选择地关注重点区域，加速视觉处理过程，完成视觉任务。本节介绍视觉显著性模型，模拟人类视觉注意机制，提取场景的显著性度量。

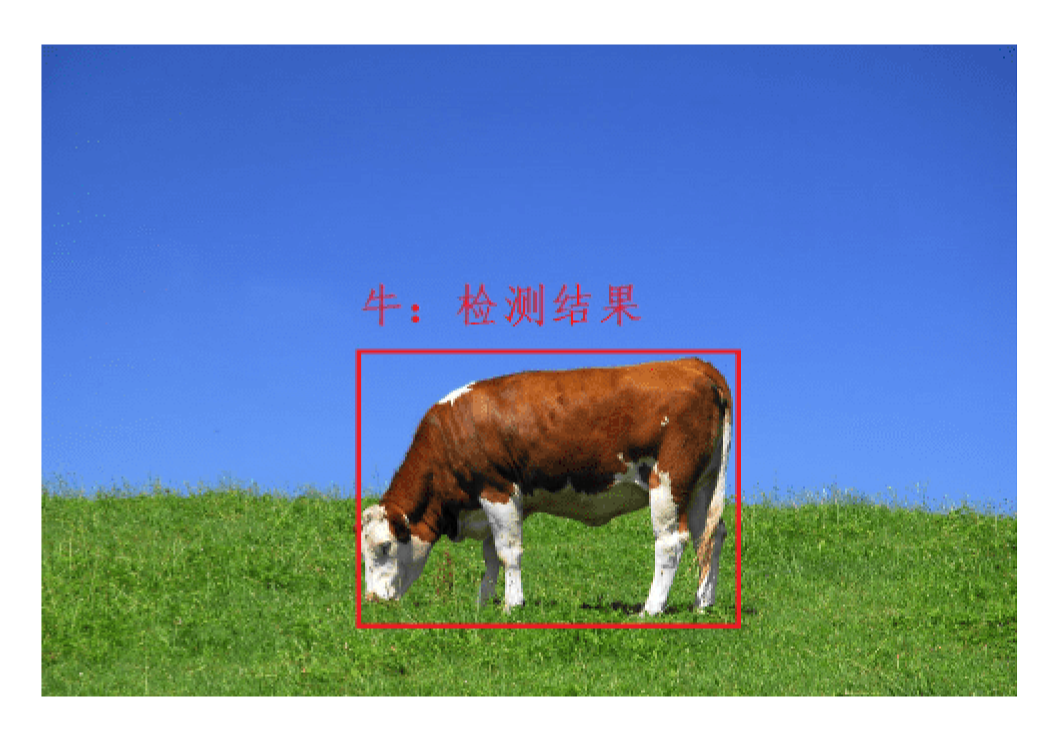
1. 视觉显著性模型简介

视觉显著性模型可以分为自顶向下和自底向上两种思路。自顶向下是任务驱动模型，针对特定目标，其实现与具体任务相关，速度较慢。自底向上是特征驱动模型，不针对特定目标，具有一定通用性，速度较快。

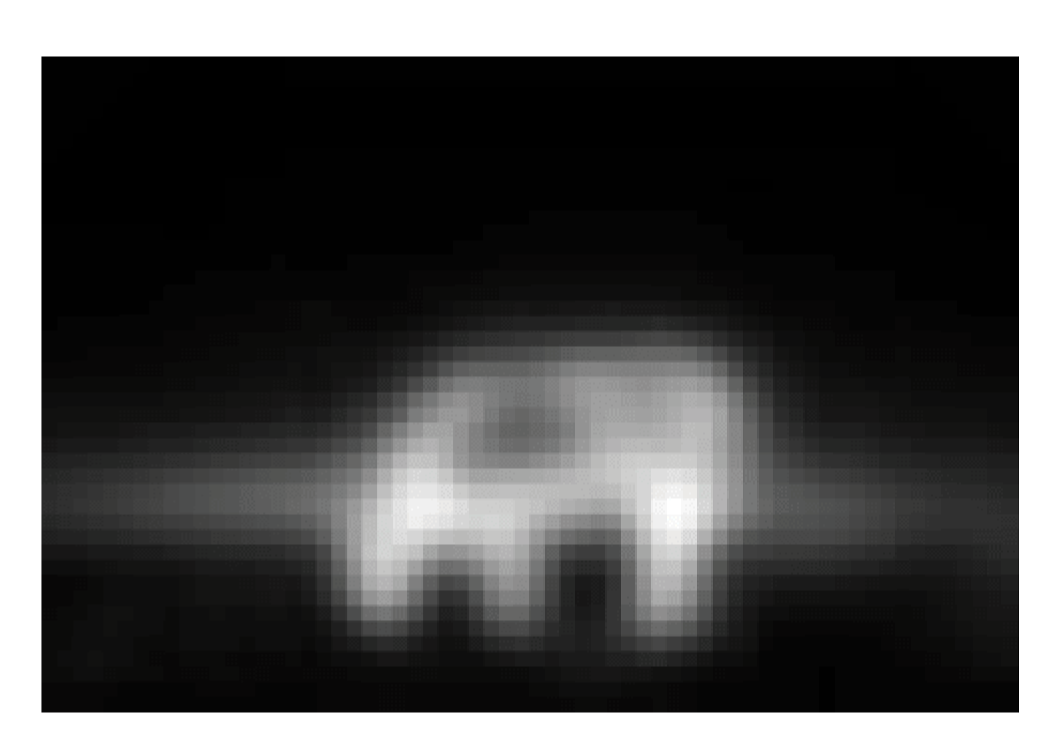
通过视觉显著性度量，有选择地关注、处理视觉信息，可以有效利用计算资源、加速算法、提高效率。

图 2.20（a）为自顶向下显著性分析示例，根据任务目标，在图像中检测牛，可以将目标检测结果的位置和置信度作为其显著性的度量。

图 2.20（b）为自底向上显著性分析示例，通过分析每个位置特征与其周边区域特征的差异性，将不同的、显著的地方凸显出来。







（a）自顶向下：目标、检测　　（b）自底向上：特征、凸显

图 2.20　视觉显著性模型示例

2. Itti 模型

Itti 模型属于自底向上的显著性模型，该模型分为 3 步：底层特征提取、显著性综合度量、注意机制模拟。Itti 模型的处理流程图如图 2.21 所示。

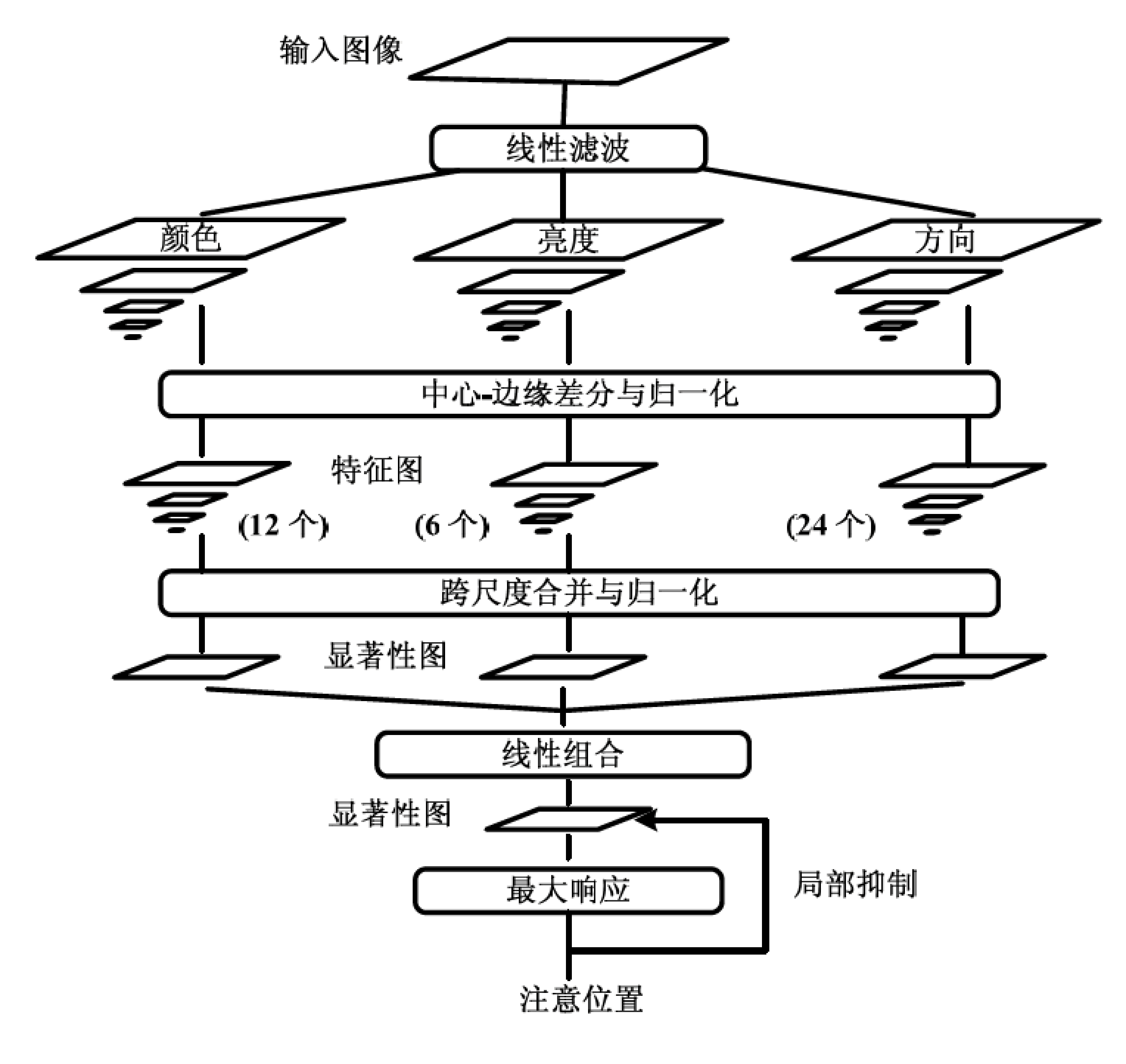


图 2.21　Itti 模型处理流程图

❑ 底层特征提取

底层特征提取包括两部分：视觉特征提取和视觉特征差分。

视觉特征提取从输入图像中提取基本的视觉特征。该模型使用3种常用的视觉特征：颜色、亮度、梯度方向。其中颜色分量建立红色R、绿色G、蓝色B、黄色Y各个分量的高斯金字塔。亮度分量使用颜色的加权平均作为亮度，并建立其高斯金字塔。梯度方向利用Gabor滤波器组实现4个方向的边缘强度计算，建立各自的高斯金字塔。

视觉特征差分利用各个高斯金字塔中不同尺度图像间的差分，得到不同位置的中心-邻域响应。在差分运算时，将低分辨率的图像插值后与高分辨率图像直接做差分。

❑ 显著性综合度量

显著性综合度量包括两部分：特征归一化和特征融合。

由于不同特征各自的特性不同，无法将其直接组合，需要经过归一化操作。归一化通过分析各个特征图的全局最大值和局部极大值，对各个特征赋予不同的权值系数。

在特征归一化之后，首先将不同尺度下的特征图统一到中间尺度，然后通过逐点相加得到各分量的显著性响应，最后将各个分量的平均值作为最终的显著性度量。

❑ 注意机制模拟

注意机制模拟可以理解为极大值注意、局部抑制、注意转移。首先关注显著性最大的点，然后抑制最大值点及其局部邻域的显著性，最后关注显著性最大的点，并以此类推。

3. Itti模型应用示例

在视频图像处理中，合理使用视觉显著性模型，可以实现高效的分析处理。假设需要设计某景区监控机器人的视觉系统，通过视觉处理算法检测场景中的行人。如果平台是持续运动的，那么背景建模方法会有一定的困难。下面给出Itti模型在该任务中的应用。

图2.22所示的3个场景，分别对应景区常见的雪地、沙漠和丘陵。为了检测其中的行人，可以使用通用的行人检测算法；为了提高检测的效率，可以融合显著性模型。3幅图像对应的显著性度量如右图所示，对于第一、第二个场景，由于人与背景特征的差异，其人所在区域显著性较大，利用显著性度量有利于行人的检测与定位。在第三幅图像中，行人区域有一定的显著性，但是由于树木和草地的区别性，树木区域也被赋予很高的显著性。在这种情况下，将显著性度量作为预处理环节，对行人检测依然有一定的帮助。

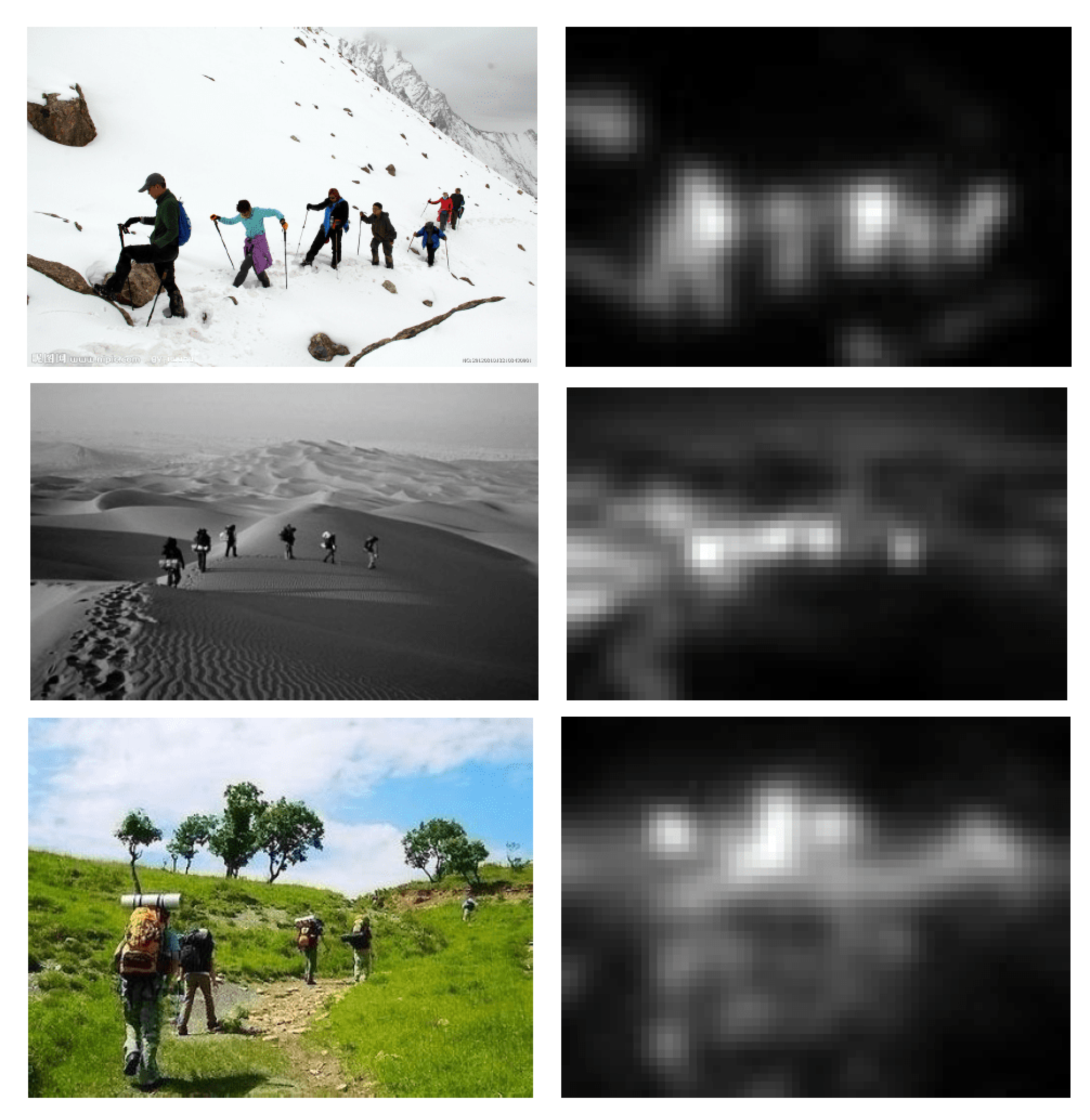


图 2.22　Itti 模型应用示例

2.9　多分辨率模型

视频图像的多分辨率模型是视频图像处理的重要方法，本节阐述多分辨率模型的背景、组成和应用。

1. 多分辨率模型的背景

在视频处理中，由于视频采集视角、焦距的多样性，目标在视频图像中往往跨越不同尺度而存在。如图 2.23 所示，不同距离上的车辆表现为大小不一的图像区域。为了可靠地检测、跟踪车辆，就需要多分辨率模型。

关于多分辨率车辆检测问题，首先建立车辆的单分辨率模型和待检测图像的多分辨率模型；然后对每个分辨率下的图像进行单独检测；最后融合不同分辨率下的输出，得到最终的多分辨率检测结果。



图 2.23 不同尺度的车辆

2. 多分辨率模型的示例

图像金字塔是最常见的多分辨率模型，有高斯金字塔和拉普拉斯金字塔两种实现形式。

❑ 高斯金字塔

如图 2.24 所示，高斯金字塔包括两步，即高斯低通滤波和欠采样。首先利用高斯核对图像进行卷积；然后欠采样，得到不同尺度下的目标图像。

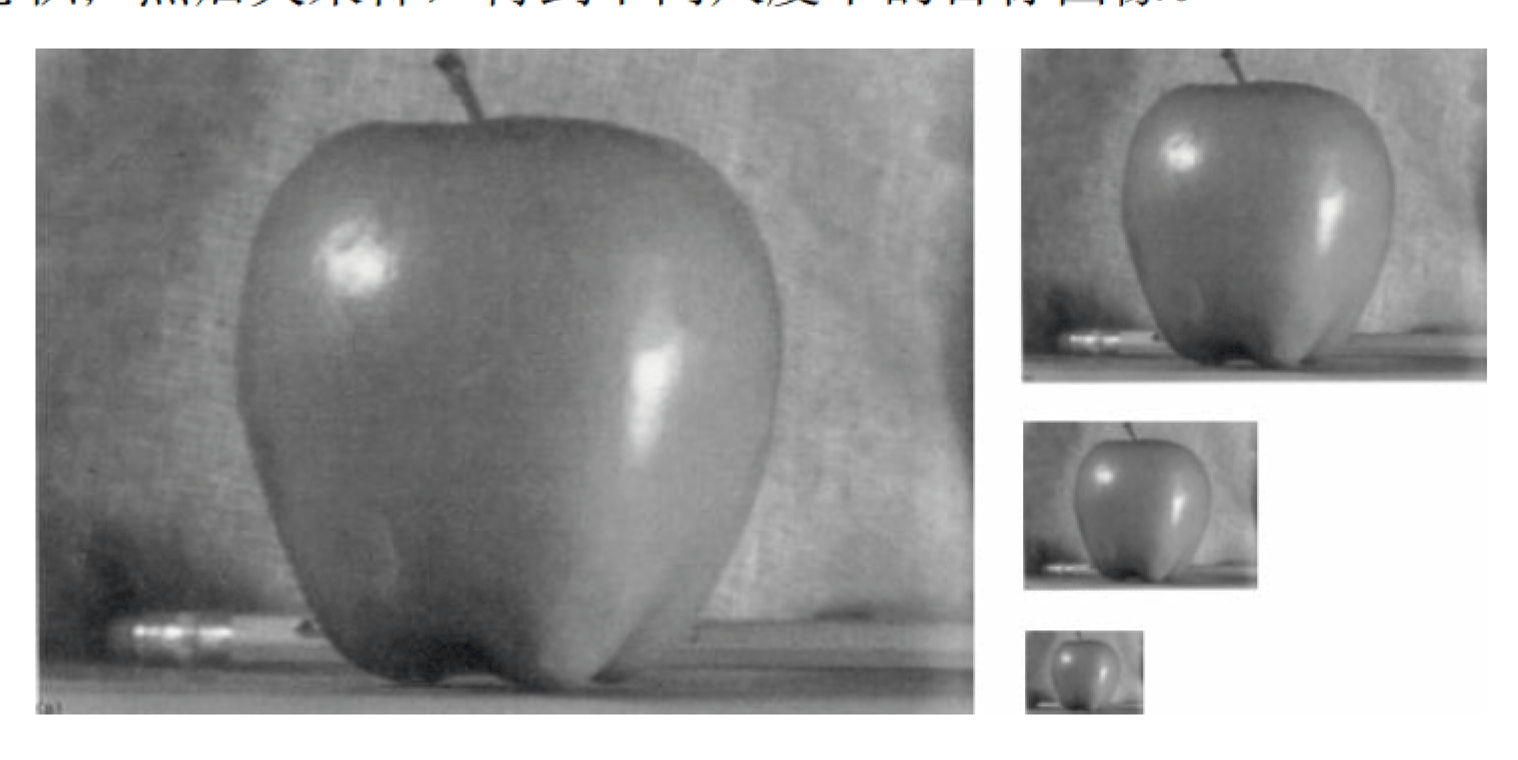


图 2.24 高斯金字塔示意图

❑ 拉普拉斯金字塔

拉普拉斯金字塔建立在高斯金字塔的基础之上，拉普拉斯金字塔就是高斯金字塔不同层之间的差分，最高层的结果两者是一样的。图 2.25 给出了与图 2.24 对应的拉普拉斯金字塔。

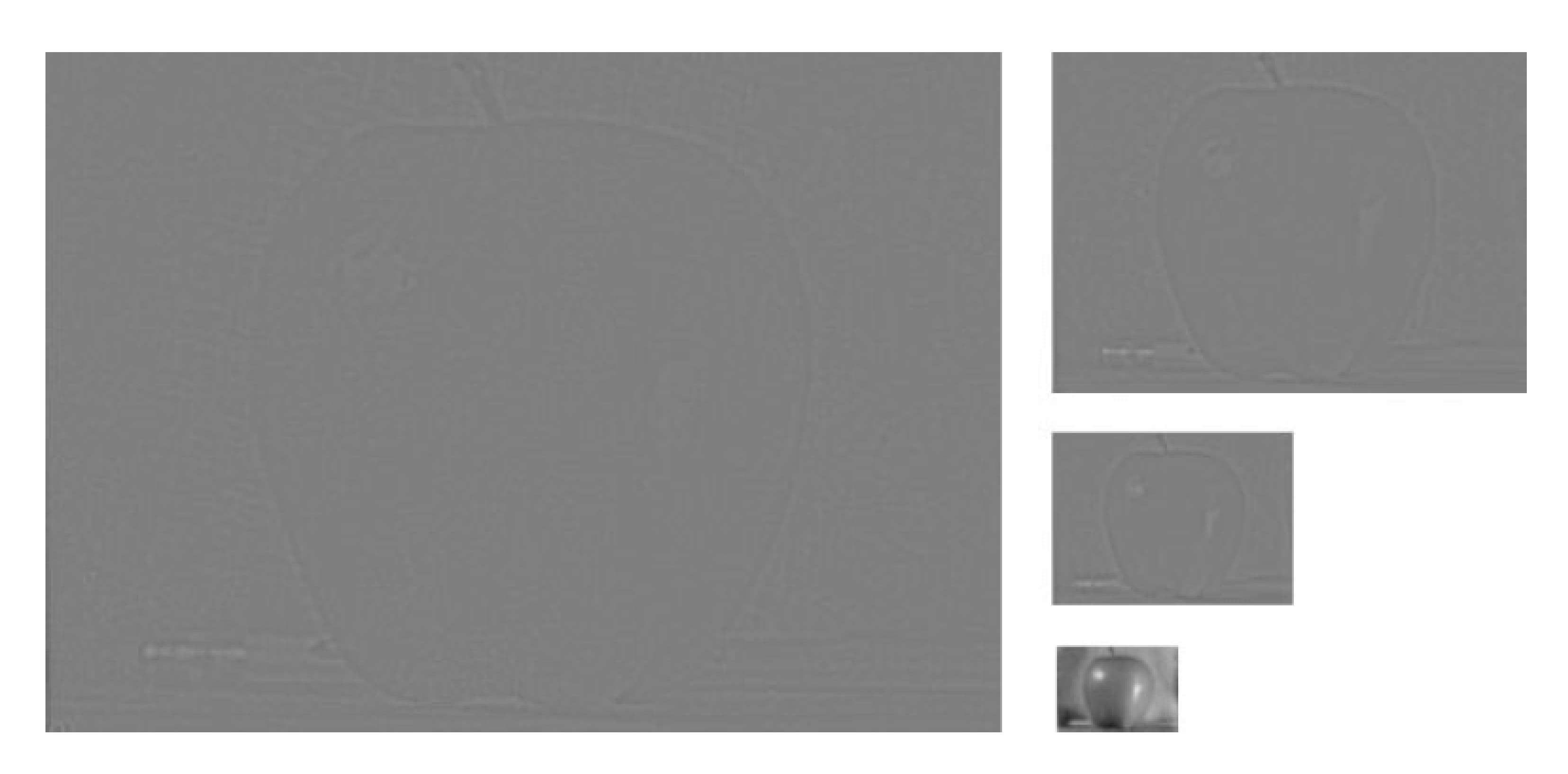


图 2.25　拉普拉斯金字塔示意图

拉普拉斯金字塔和高斯金字塔两者之间存在转换关系，假设高斯金字塔为 $(G_0,G_1,\cdots,G_n)$，拉普拉斯金字塔为 $(L_0,L_1,\cdots,L_n)$，那么有 $Li=G_i-\widehat{G}_i$。其中，$\widehat{G}_i$ 是 G_{i+1} 的扩展图像，就是通过线性插值得到的与 G_i 同分辨率的图像。

3. 多分辨率模型的应用

高斯金字塔常用于图像多尺度分析，如在不同尺度下的目标检测；而拉普拉斯金字塔常作为图像压缩、图像降噪的基础，如对特定尺度下的细节进行平滑和增强。

以拉普拉斯金字塔用于图像细节增强为例，首先将原始图像分解为两层的拉普拉斯金字塔；然后将第一层的高频分量增强（乘以 4），再与第二层图像一起重建原始图像，可以得到增强后的效果图。图 2.26（a）、（b）分别对应增强前后的图像。



（a）原始图像



（b）增强的图像

图 2.26　基于拉普拉斯金字塔的图像细节增强

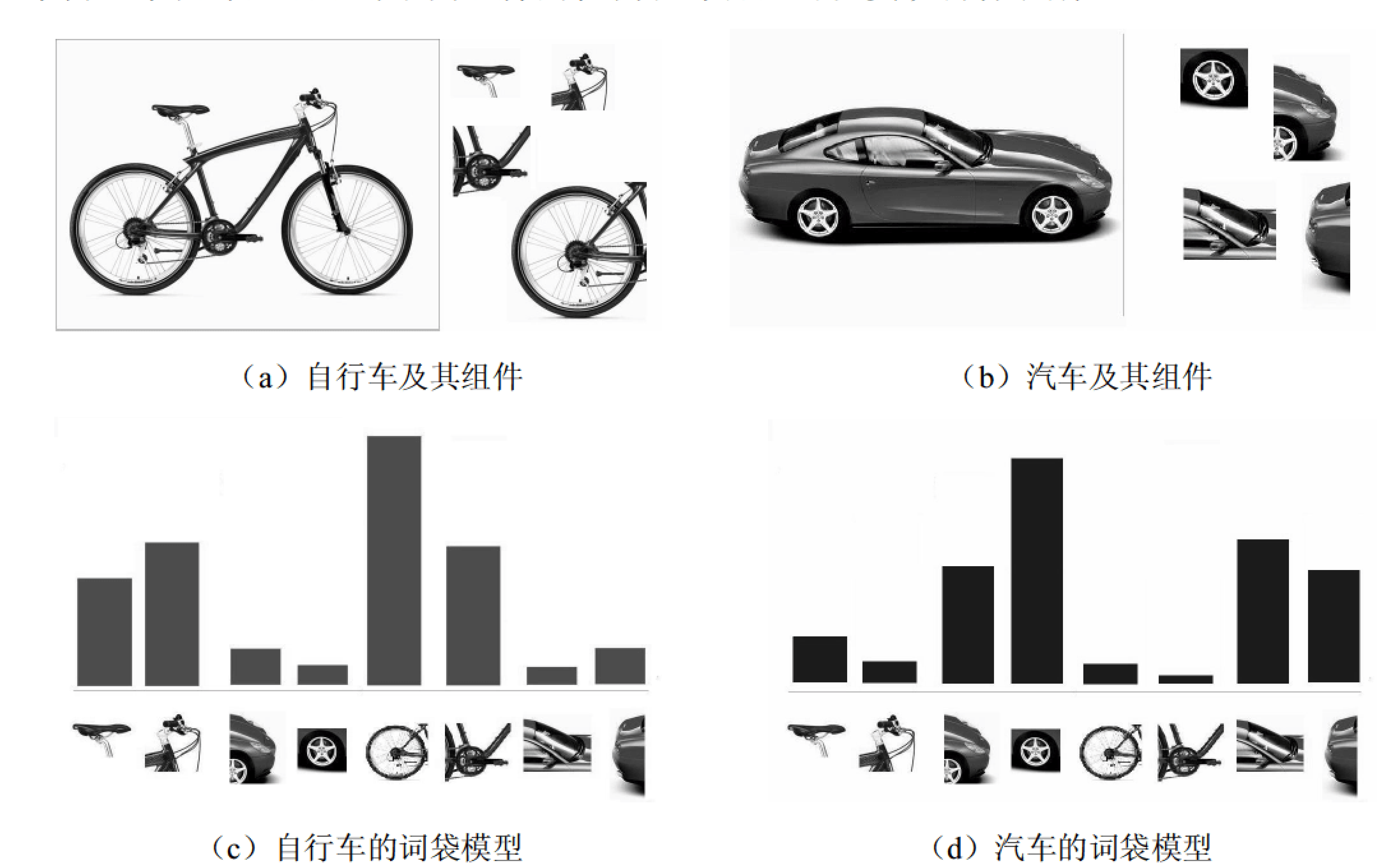
2.10　视觉词袋模型

视觉词袋模型从文本分析方法借鉴而来，广泛应用于视频图像处理，如目标识别。

1. 视觉词袋模型的原理

视觉词袋模型来源于文本分析，将文章理解为词语的集合，利用文章中词语的直方图分布来表述文本，实现文本识别。将文章、词语推广到视觉中的目标、组件，得到视觉词袋模型。

如图 2.27 所示，自行车和汽车都可以表示为不同组件的集合，如把手、轮胎、坐垫、车窗、车轮等，通过对不同组件的检测和综合，可以实现目标识别。



（a）自行车及其组件　　（b）汽车及其组件

（c）自行车的词袋模型　　（d）汽车的词袋模型

图 2.27　自行车和汽车的目标与组件

2. 视觉词袋模型的建模

视觉词袋模型的建模包括 3 个步骤：特征定位与描述、字典构建、目标表示与分类。

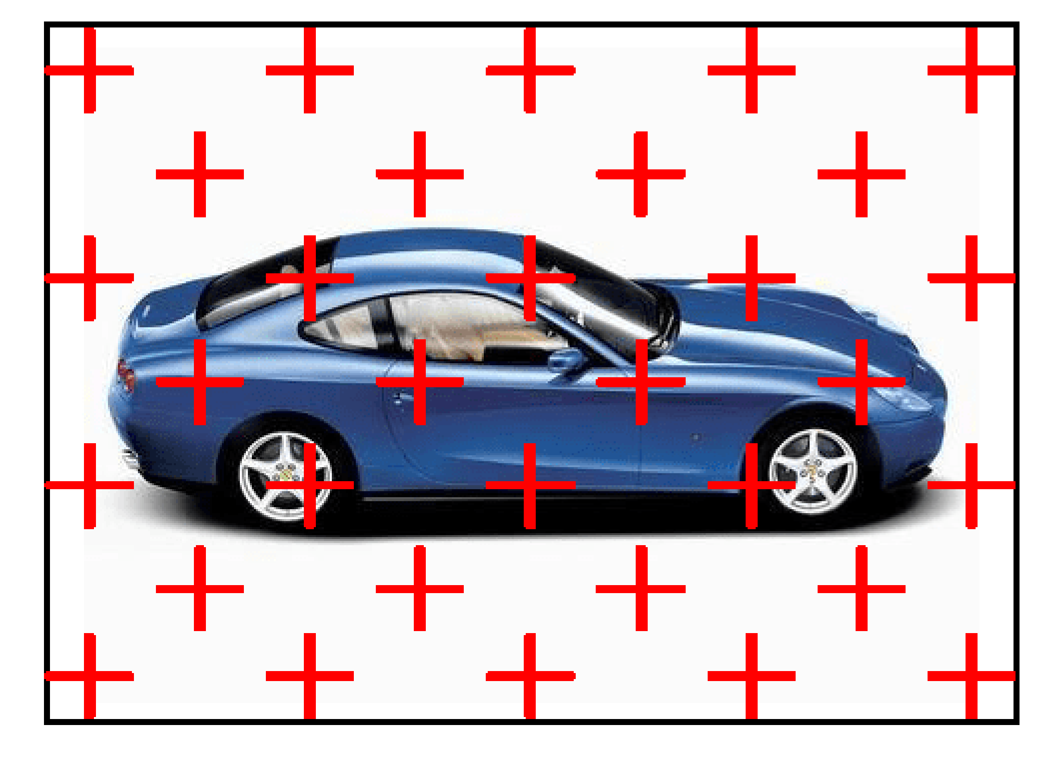
❑ 特征定位与描述

特征求取包括两个关键步骤：特征定位和特征描述。

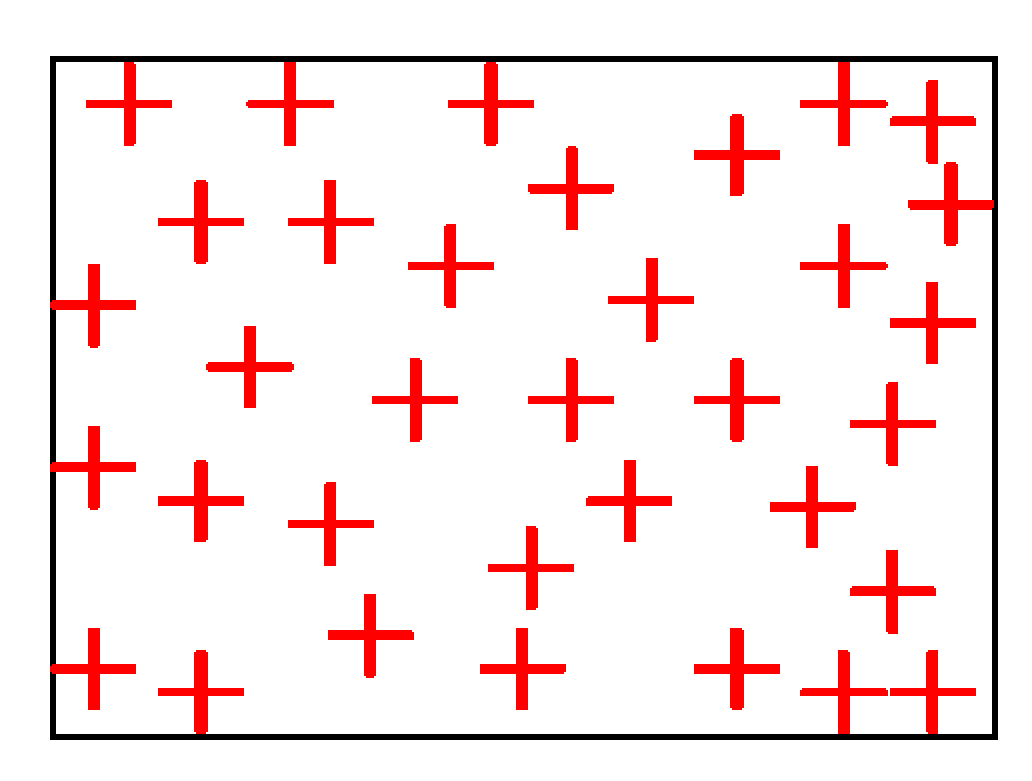
常用的特征定位方法有规则采样、随机采样、关键点检测、基于分割的方法等，图

2.28 给出了 4 类方法的示意图。前 3 类先得到点的位置，然后计算点及其邻域的特征描述；基于分割的方法，将每个分割出的小块作为计算特征描述的单元。

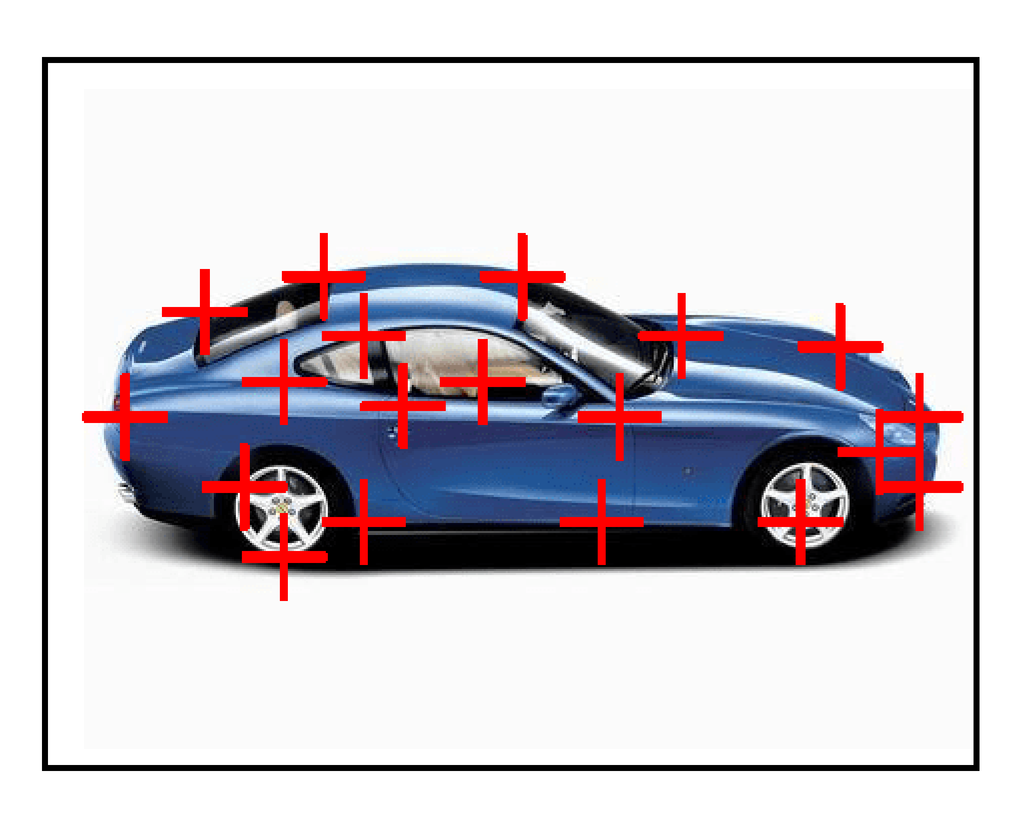
常用的特征描述方法有：颜色、梯度方向、区域统计等。



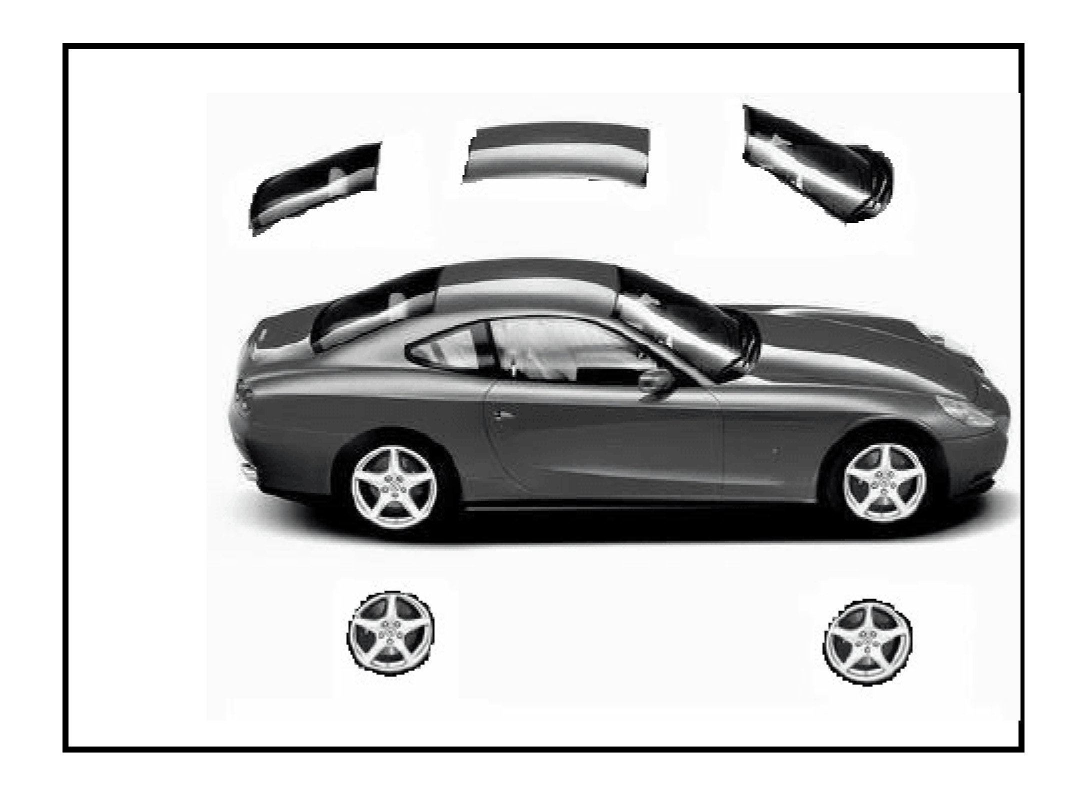
（a）规则采样



（b）随机采样



（c）关键点检测



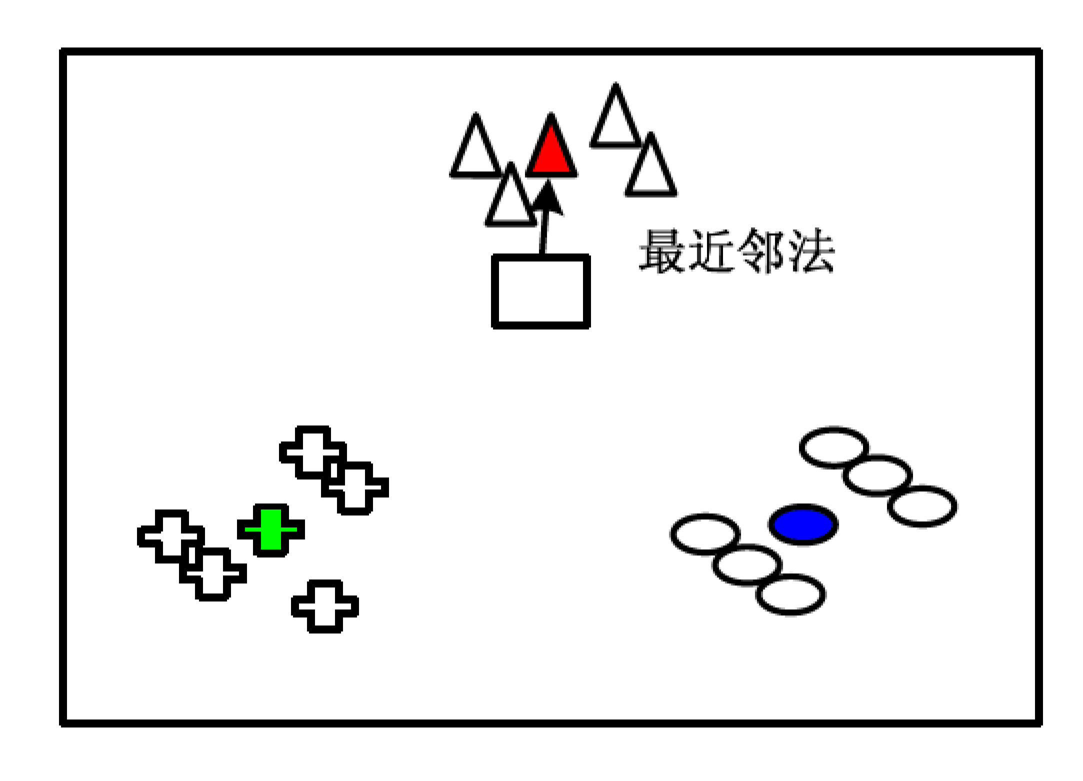
（d）基于分割的方法

图 2.28　特征定位方法

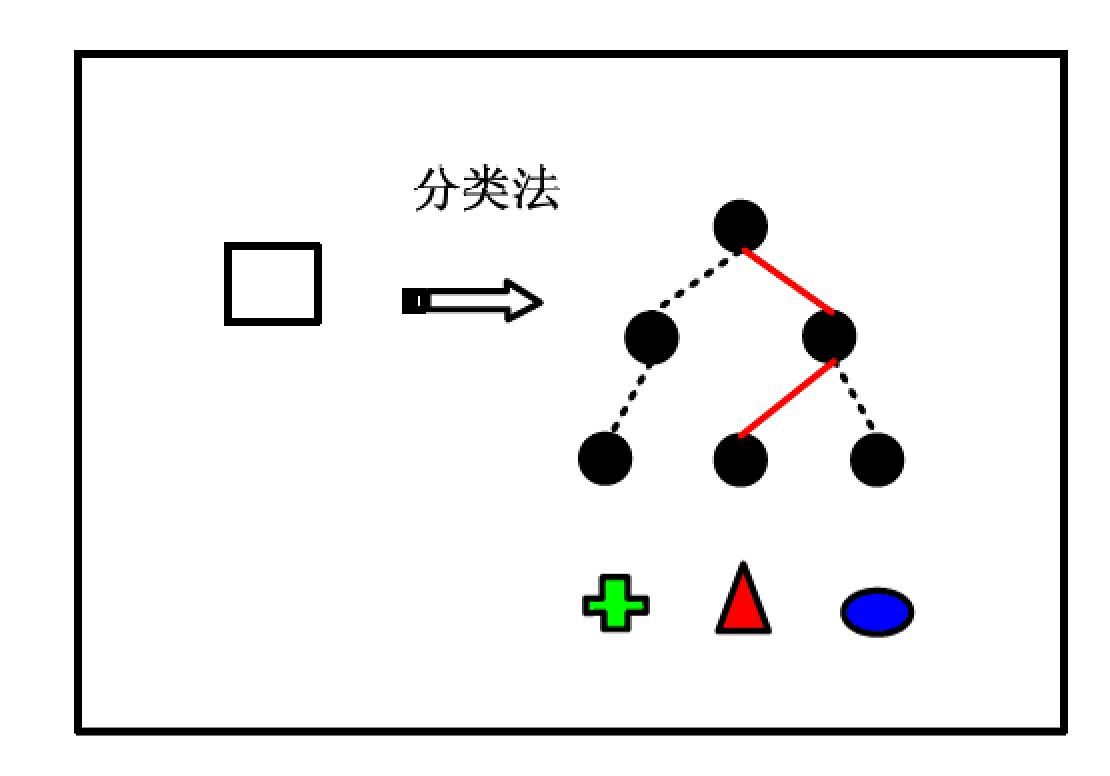
❑ 字典构建

利用训练集中的特征建立字典，然后采用字典表示各个特征。字典构建有多种方法，可以是基于无监督的聚类，如 K-Means；也可以是基于有监督的分类，如 Random Forest。

如图 2.29 所示，在建立字典之后，将各个特征用字典来表示。可以用最近邻表示法，每个特征用其在字典中最接近的一个条目表示。也可以用基于有监督的分类方法，根据分类结果所在的分支确定其归属。



（a）聚类-最近邻方法



（b）分类方法

图 2.29　字典的构建

❑ 目标表示与分类

在建立字典和利用字典表示各个特征之后，统计得到图像区域总体的直方图分布。归一化之后的直方图作为图像的整体特征，用于分类和识别。在分类和识别中，常用的是支持向量机（SVM）。

3. 视觉词袋模型的讨论

视觉词袋模型通过构建字典，将视觉特征转化为直方图，以此进行分类和识别。该算法通用性较好，对目标的位移、旋转等姿态变化有一定的鲁棒性，应用广泛。

如图 2.30 所示，视觉词袋模型没有使用特征之间的相对位置信息，目标不仅是由组件构成的，组件的相对位置关系也是构成目标的要素。利用组件及其相对关系，有助于提升目标识别和定位的效果。



图 2.30　目标 ＝ 组件 ＋ 相对位置

2.11　视频语义模型

随着视频监控设备的广泛使用，视频数据大量涌现。如何合理地管理、利用海量的视频数据，已经成为当前研究的重点和难点。有效利用视频数据的核心就是高效、精确的视频信息检索。本节介绍视频语义模型及其在视频语义检索中的应用。

1. 视频语义模型的简介

数字视频技术将视频数据编码为视频流，使计算机可以采用数学方法表示、存储和处理视频数据，极大推动视频处理技术的发展。然而，这样的表示方法依然无法实现高效的语义理解和检索。

如图 2.31 所示，视频语义模型将视频数据中的对象、事件及其相互关系有效组织起来，为视频语义检索提供支持。



（a）初始帧



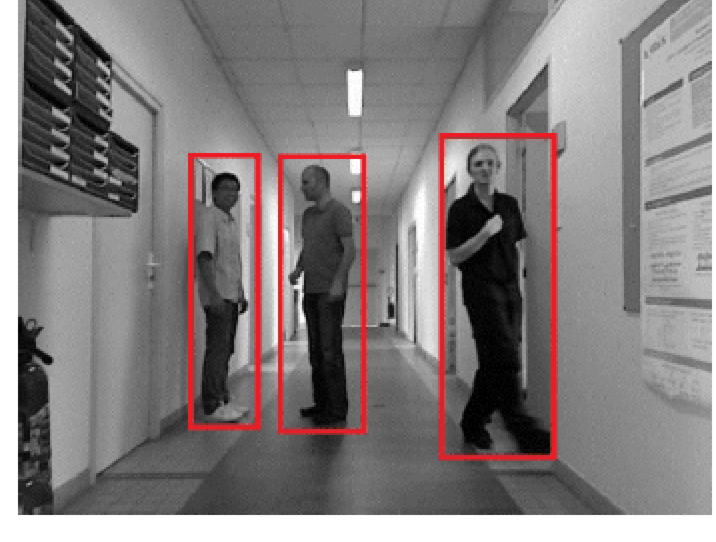
（b）中间帧



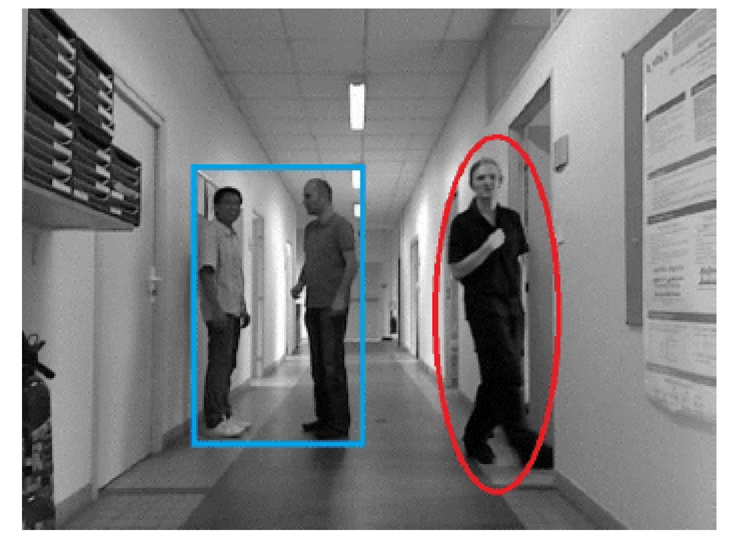
（c）结束帧



（d）前后帧的差分



（e）人体检测结果



（f）行为检测结果

图 2.31　视频语义检索示意图

图 2.31（a）、（b）、（c）是一个视频流中的三帧图像，各帧间隔 2s 左右。在该视频片段中，有 3 个人物，其中两人在交谈，另外一人完成一个进门动作。

针对该视频片段，假定 3 个不同层次的检测任务：运动目标检测、人体的检测、行为检测。运动目标检测可以通过背景建模、图像差分实现，是对底层特征的操作；人体

的检测可以通过分析单帧图像实现；行为检测需要通过对多帧图像的分析得到。视频语义检索，就对应诸如“找出所有两两交谈的人”、“找出所有进入某房间的人”的处理任务。

为了实现视频语义检索，必须建立视频语义模型。在该任务中，需要包含对象（如人、门）、事件（如交谈、进门）和关系（空间位置、时间先后）。交谈可以通过检测人及其相互关系来实现，两个人在空间上邻近、姿态是面对面，甚至伴有一定的肢体语言，可以初步判断为交谈行为。进门可以通过检测人、门及其相互关系来实现，一个人开始在门外，然后在门中，之后消失在门后，可以认定为进门动作。

2. 视频语义建模与检索

视频语义模型需要分层结构，视频语义模型的构建与具体需求相关，需要根据实际情况设计其功能和复杂度。

如图 2.32 所示，4 层的视频语义模型如下：

- 第一层，原始数据，逐帧存储。
- 第二层，底层特征，如运动目标、光流、颜色、纹理。
- 第三层，中层信息，包括目标识别结果、场景及空间关系。
- 第四层，高层语义，包括各种行为的识别结果。

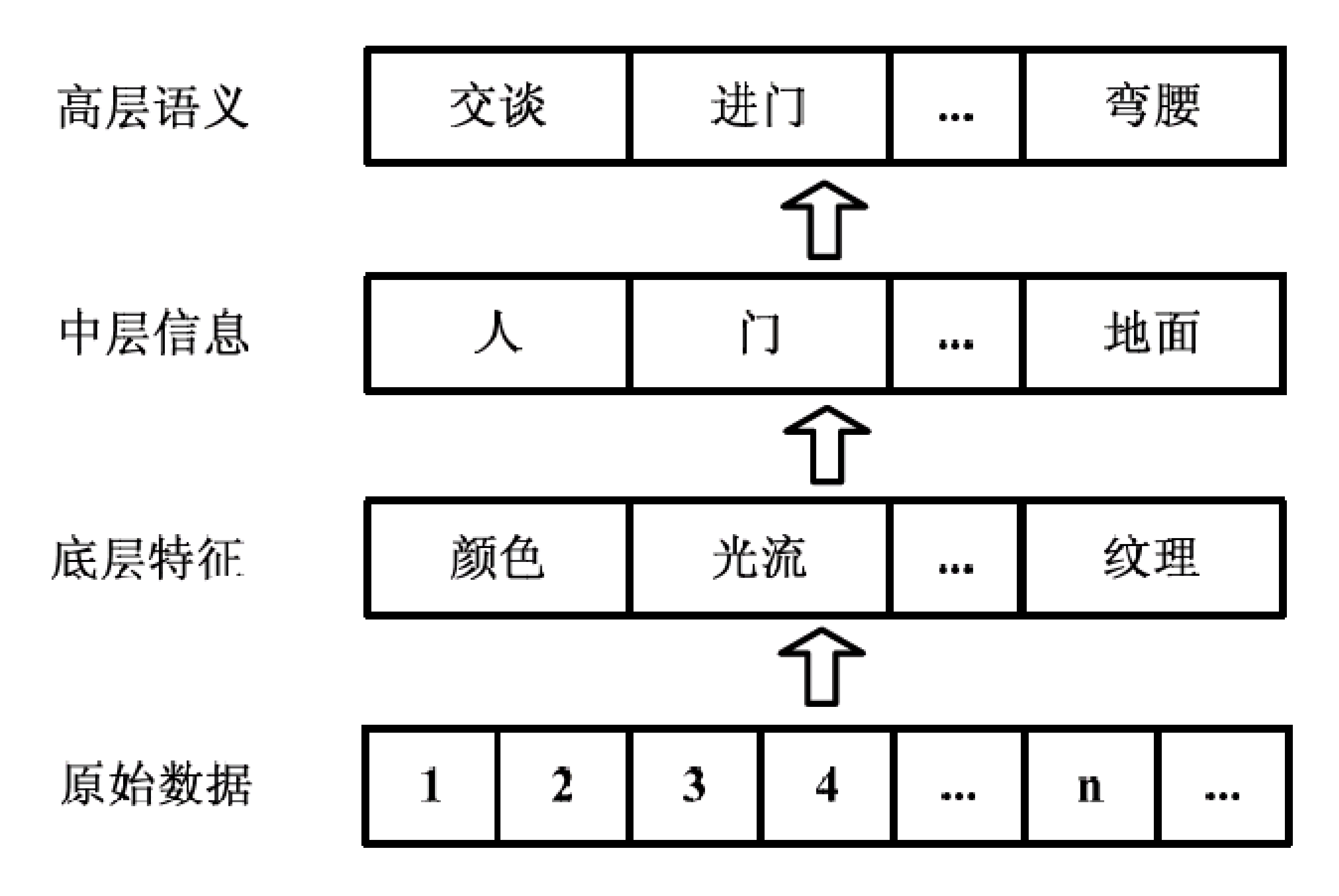


图 2.32 视频语义模型

从原始数据到高层语义的构建，即视频语义建模，涉及图像处理、模式识别、数据库等技术。图像处理支持光流计算、背景建模；模式识别实现目标检测；数据库技术可以有效组织数据，高效检索。

如图 2.33 所示，从高层语义到原始数据的查询，即视频语义检索。如查询“所有的

两两交谈”，从高层语义中提取交谈行为，对应到中层信息中的交谈的人，以及相应底层特征的颜色、纹理及对应的原始帧号。输出的结果可以是：在第 m 帧到第 n 帧期间，人物 A 与人物 B 进行交谈，人物 A 穿着白色格子花纹上衣，人物 B 穿着灰色上衣。

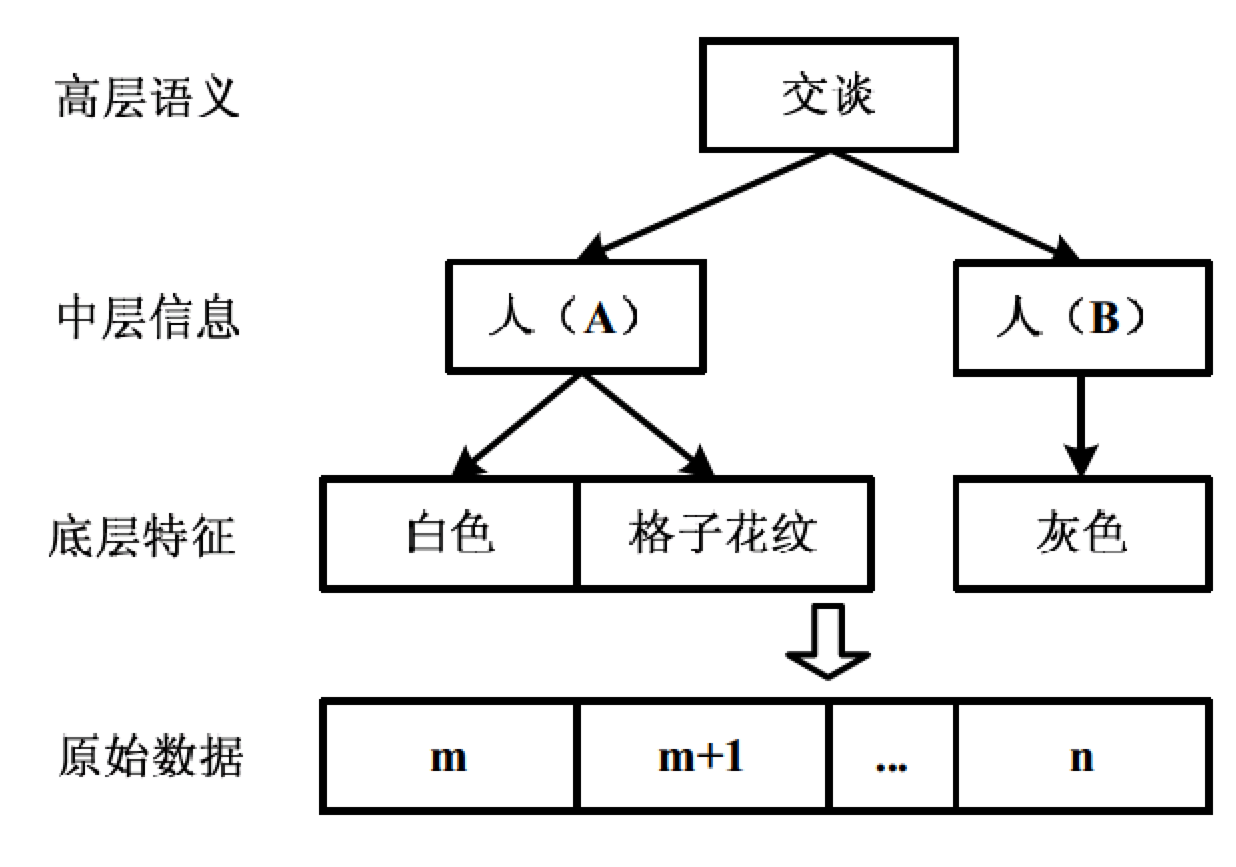


图 2.33　视频语义检索

3. 视频语义模型的讨论

视频语义模型的设计与使用包括 3 个关键点：模型的语义表达、模型的语义获取与分析、模型的语义查询。

- 模型的语义表达

需要支持的对象、事件和关系；需要支持的对象、事件和关系的属性；是否支持约束；是否支持概率推理等。

- 模型的语义获取与分析

是否集成领域知识，如何集成；是否需要人工标注，是否涉及人机交互；是否可以推导隐含信息；是否可以检查逻辑错误等。

- 模型的语义查询

需要支持的查询种类和接口，是否支持增量查询，是否支持推理。

第 3 章

海量视频管理

随着视频监控系统和网络视频的广泛应用，海量视频数据急剧膨胀。视频信息数据量大，抽象程度低，导致处理能力不足，大量视频数据不能得到有效利用。面向 PB 级以上的海量视频管理成为研究热点，核心难题集中在海量视频的存储架构、管理模型、数据库和管理系统等方面。

3.1　视频数据库

3.1.1　海量视频数据

公共视频监控系统、个人视频采集设备以及互联网的迅猛发展，极大地方便和丰富了人们的生活、学习和工作，改变了人们的交流方式。

为了解决视频信息膨胀问题，对包含大量非结构化信息的海量视频数据进行组织、表达、管理、查询和检索成为迫切需求。

海量视频数据形式多样、类型各异，具有如下特点。

1. 数据量巨大

数秒钟的视频片段其存储空间可能为几兆字节，将对数据库的组织和存储方法产生影响。

对能处理连续数据的视频数据库要求具有高速性能。

2. 存储方式多样化

无法将所有的视频信息保存在某台设备上，常用网络分发，对视频库的数据存取构成挑战。

3. 媒体特性差异大

媒体种类的增多增加了数据处理的复杂度，视频不仅具有多种分辨率、视场、大小，而且视频文件有多种存储格式，如 AVI、MPEG-X、MJPEG、H.26x、ASF、RM 等。

不同格式、不同类型的视频文件其数据处理方法各不相同，因此需要视频数据库具有不断增加新的媒体支持类型及相应处理方法的能力。

4. 接口形式复杂

视频数据具有复合、分散和时序等特性，采用简单的基于字符的检索方式效果较差，而应采用基于视频内容语义的检索方式。

视频对数据库的影响涉及数据库的用户接口、数据模型、体系结构、数据操纵以及数据应用等方面。

3.1.2　面向对象的海量视频数据库

1. 传统数据库的局限性

传统数据库主要依赖人工分析实现视频标注，建立类似于文本文献的索引数据库，通过检索获得视频编号，利用这些编号获取对应视频，属于关系数据库。在海量视频背景下，传统数据库的局限性有：

- 对视频加注文本信息由手工完成，费时费力；
- 文本描述信息是操作者的主观描述，导致描述多样化；
- 文字标注难以刻画视频的全部内容；
- 文字标注具有语言、民族、地域差异，难以成为通用描述。

为了克服传统关系数据库工作量大、主观性强和特征描述能力有限的弊端，首先由计算机自动提取视频对象的高层次特征，然后进行视频分割，最后按照这些客观特征进行大规模的视频数据检索。

2. 面向对象的视频数据库的特点

面向对象的视频数据库的结构如图 3.1 所示。

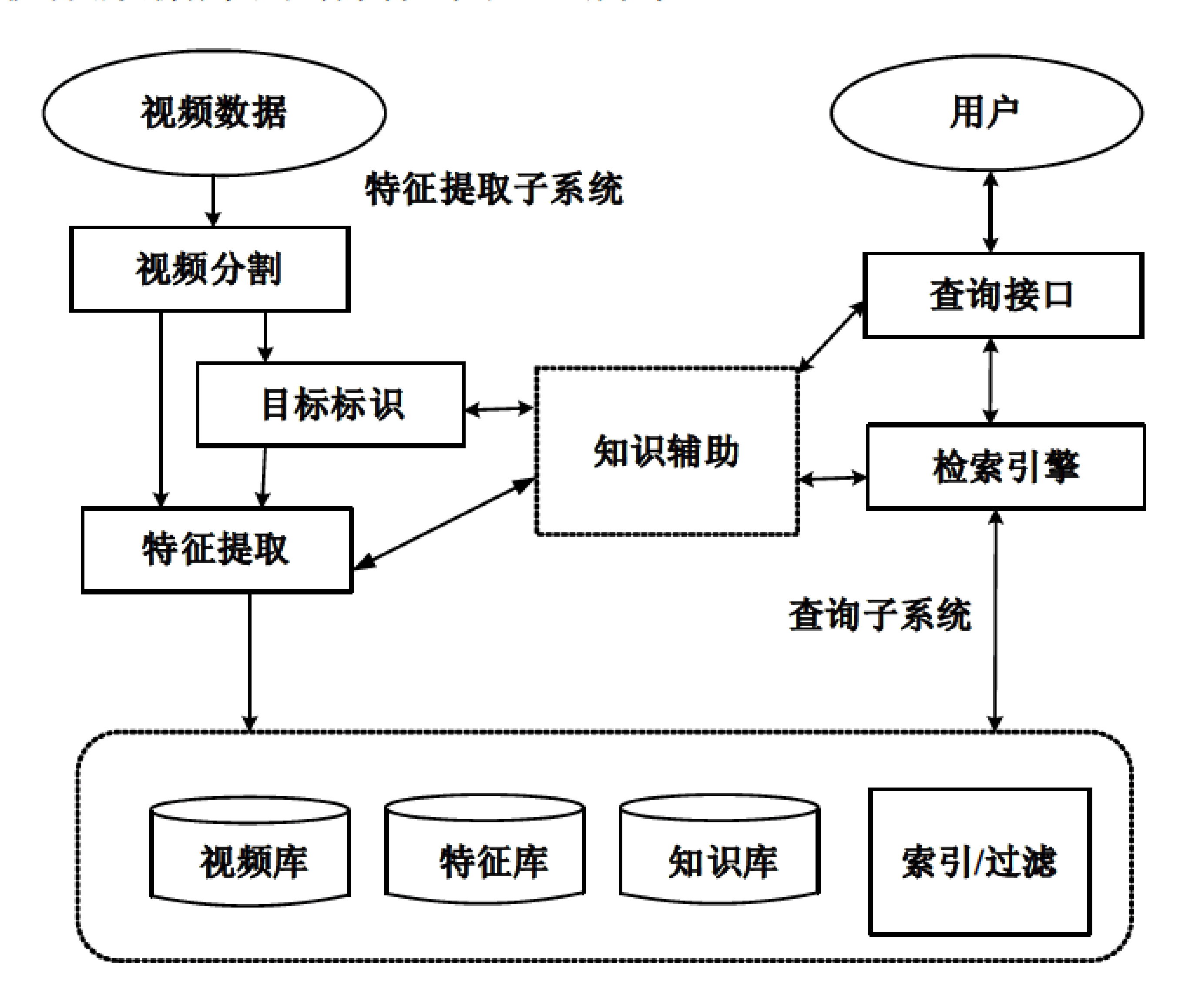


图 3.1 基于特征自动提取的视频数据库

面向对象的视频数据库突破了传统数据库的局限性，融合了模式识别、计算机视觉、图像理解等技术，具有 5 个显著特点：

- 直接分析图像内容，提取语义级特征，检索更有效，适应性更强；
- 字符检索采用精确匹配方式，图像检索采用相似匹配方式；
- 由用户参与的检索过程，可以不断改进检索方式，交互性强；
- 包括图像库、特征库和知识库，可以满足多层次的检索要求；
- 图像检索采用示例查询法，当用户不清楚准确的检索要求或图像信息时，可以输入或选择相似的示例图像，或是绘制参考图形作为检索条件，利用检索结果进行检验，对检索条件做出修正。

3.2 集中式视频数据库

如图 3.2 所示，集中式视频数据库由中心处理器、视频数据存储设备、其他外围设备组成，该数据库物理上被定义为专有位置，具备数据处理和管理能力，用户可以在相同站点，或位于其他位置的站点上通过远程终端操作。

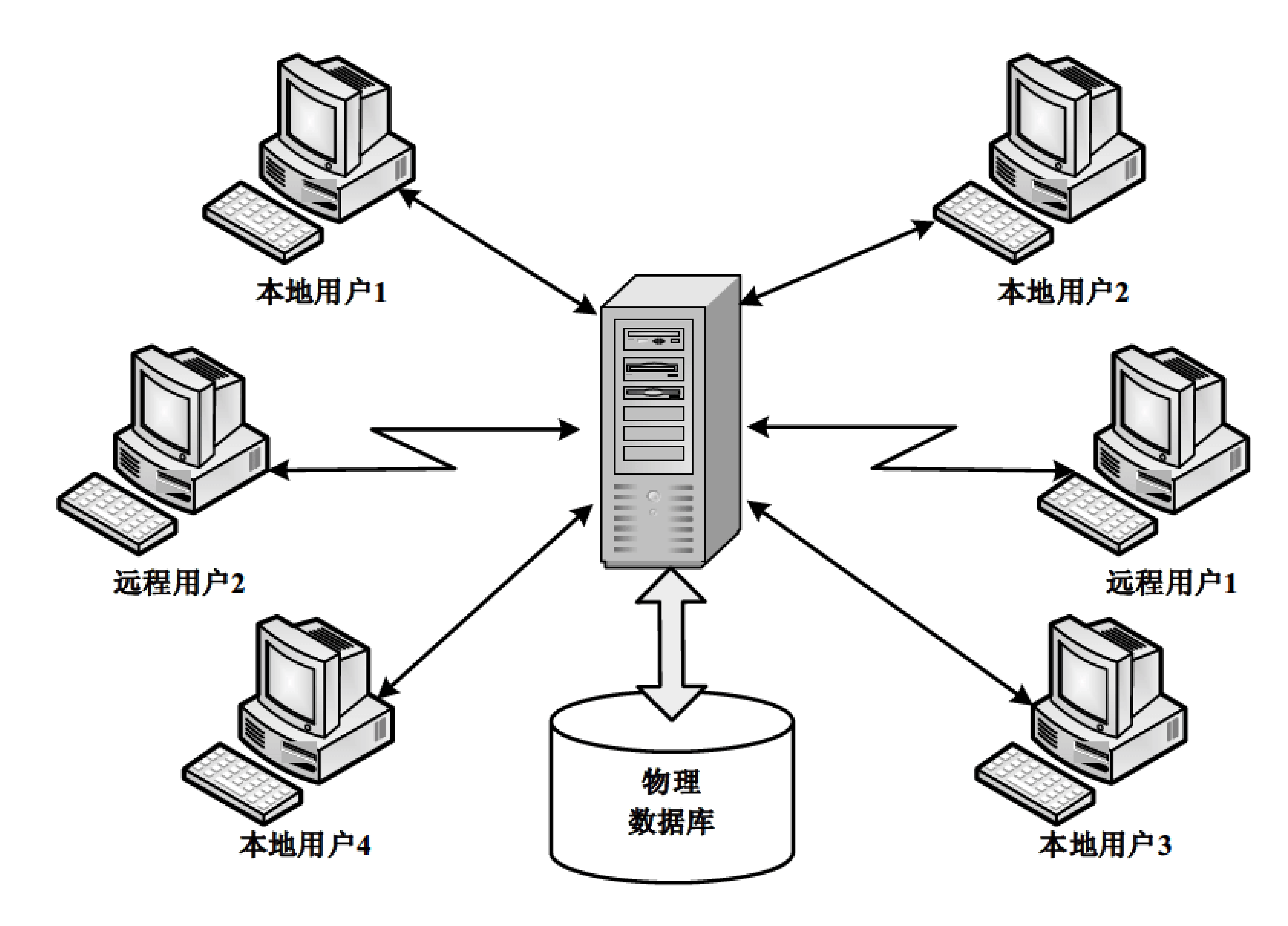


图 3.2　集中式视频数据库

1. 集中式视频数据库的优点

在集中式视频数据库里，便于实现常用的查询、编辑、备份、权限设置等大多数功能。

数据库所处位置灵活，小型用户可以在 PC 上建立数据库，大型用户可以由大型机控制整个数据库。

2. 集中式视频数据库的缺点

所有用户必须依赖于中心站点计算机或数据库正常运行。

从终端到中心站点的通信开销昂贵。

3.3　分布式视频数据库

随着视频监控系统和互联网的发展，海量视频数据急剧膨胀，需要安全、高效地保存、分析这些数据，分布式数据存储技术可以较好地满足此要求。利用该技术，数据被存储在物理上分散的多个节点上，此类节点资源被统一管理与分配，向用户提供访问接口，从而可以解决本地文件系统在大小、数量等方面的限制问题。

3.3.1　基于 Hadoop 的视频数据库

Hadoop 是一种分布式系统的基础架构，由 Apache 基金会开发，可用于进行大规模数据处理，具有如下特点。

- 海量存储：能可靠地存储和处理 PB 级数据。
- 成本低：常用普通机器组成的服务器集群，可达数千个节点。
- 高效率：通过分发数据，可以在数据所在节点上并行处理。
- 可靠性：能自动维护数据的多个备份。

Hadoop 主要由 HDFS、MapReduce 组成，HDFS 实现对大规模数据的分布式存储管理，而 MapReduce 则对大规模数据进行分布式计算。

1. HDFS

如图 3.3 所示，HDFS（Hadoop Distributed File System）基于 JAVA 的主/从模式，支持数据密集型分布式应用。HDFS 由一个命名节点和若干个数据节点组成，数据节点的数量不限，根据实际需求而定，可以从一个至数千个。

每个文件被分成若干数据块，这些数据块被存放到一组数据节点之上；数据节点根据命名节点的指令，对数据块进行创建、删除和复制等文件管理操作。

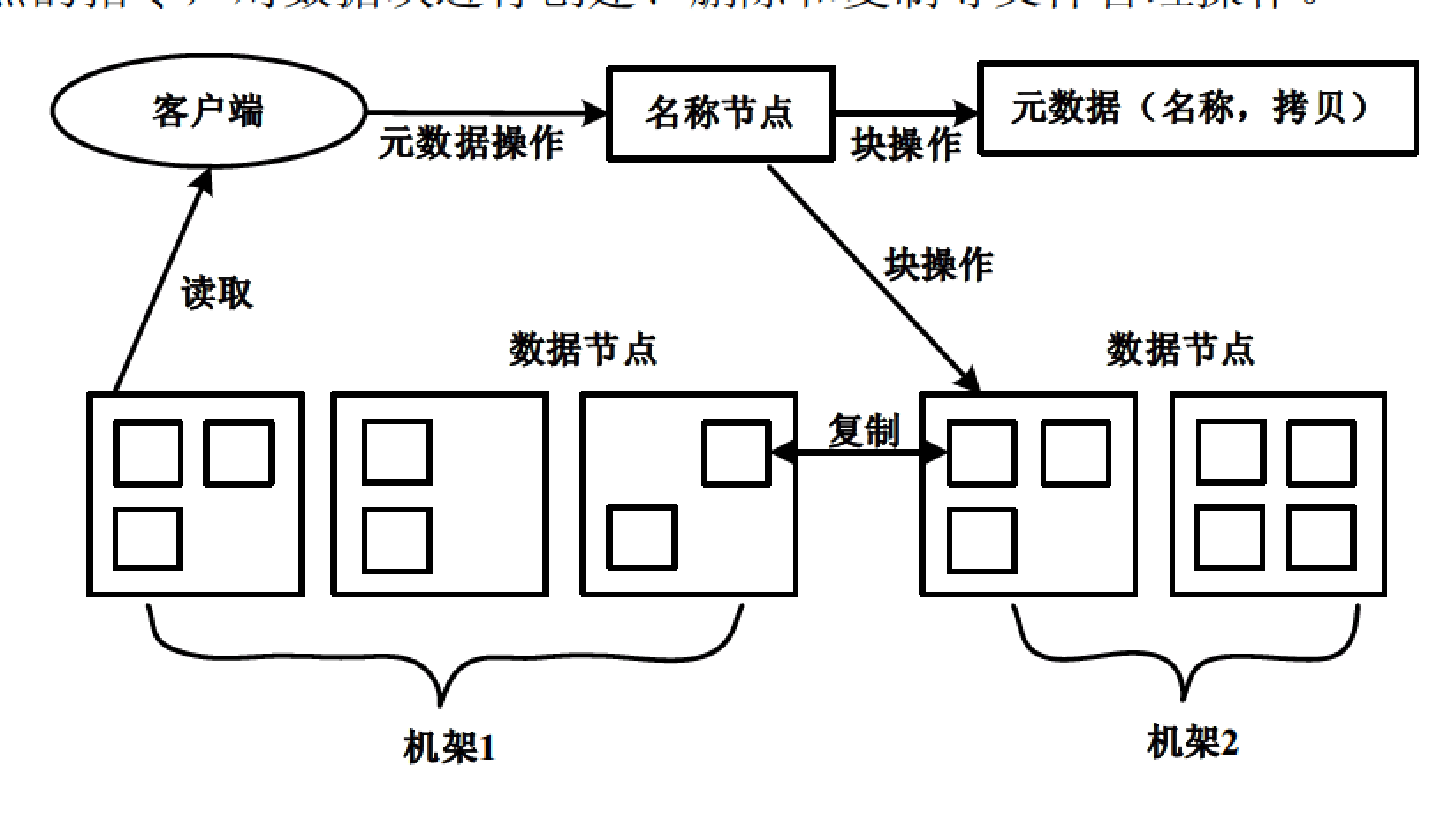


图 3.3　HDFS 结构

2. 数据组织

（1）命名节点（元数据节点）

命名节点在内存中保存文件系统的元数据信息，元数据信息包括：

- 文件列表信息；
- 每个文件的块列表；
- 每个块对应的数据节点；
- 文件属性，包括创建时间、创建者、副本份数等。

如图 3.4 所示，命名节点的文件夹包括 edits、fsimage、fstime 等文件。

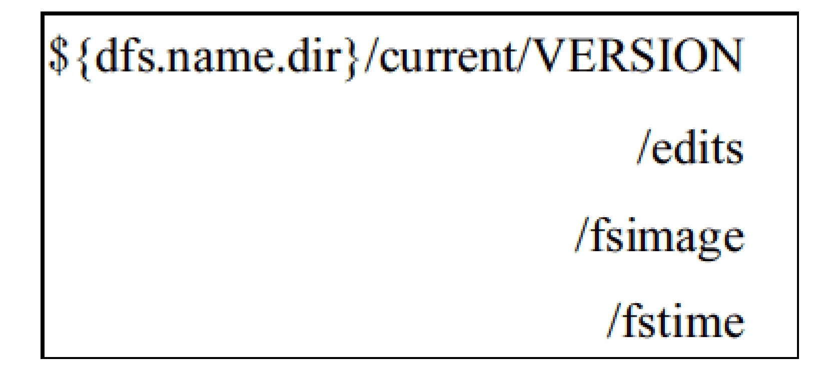


图 3.4　命名节点文件夹的结构

- edits 文件，记录文件系统的变化，如创建、删除、文件副本数等；
- fsimage 文件，是命名节点的映像文件，元数据在磁盘上的 checkpoint；
- fstime 文件，记录 checkpoint 的时间。

（2）数据节点

数据节点负责数据存储，数据块的复制操作由数据节点之间的通信完成；当在客户端写文件时，数据节点之间相互配合，以保证逻辑一致性。

如图 3.5 所示，安装 Hadoop 时，数据块存放目录由配置文件指定，数据存放在设定文件夹的 dfs/data/current 目录。

subdir61	128 项
subdir62	128 项
subdir63	128 项
bik_244953846541835520	21.6KB
blk_244953846541835520_4260.meta	183B
blk_1096531559384704428	35B
blk_1096531559384704428_1008.meta	11B
blk_1293064074033438457	139.1KB

图 3.5　current 目录结构

current 文件夹内包括子目录、数据块文件和数据块元数据文件，子目录名从 subdir0 到 subdir63，子目录下有数据块文件和数据块元数据。

数据块文件和元数据文件的实例为：

- blk_<id>，存放 HDFS 中具体的数据块；
- blk_<id>.meta，保存数据块的元数据，如版本、类型和 checksum 等。

3. 数据流

（1）读文件流程

如图 3.6 所示，从 HDFS 读文件涉及 client、NameNode、DataNode 3 个进程和 DistributedfileSystem、FSDataInputStream 等类的操作。

步骤01 client 调用其 get 方法获得 HDFS 文件系统的一个实例（DistributedfileSystem），然后调用 DistributedfileSystem 的 open 方法。

步骤02 DistributedfileSystem 通过 RPC 远程调用 NameNode，取得文件数据块的位置信息；对于每个数据块，NameNode 返回数据块所在的 DataNode（包括副本）的地址；DistributedfileSystem 返回 FSDataInputStream 给 client 用于读数据。

步骤03 client 调用 FSDataInputStream 的 read 方法。

步骤04 FSDataInputStream 连接保存此文件第一个数据块的最近数据节点，读数据块，传回给 client。

步骤05 当第一个数据块读完，FSDataInputStream 关闭与该 DataNode 的连接，然后开始读第二个数据块。

步骤06 当 client 读文件结束，调用 FSDataInputStream 的 close 方法。

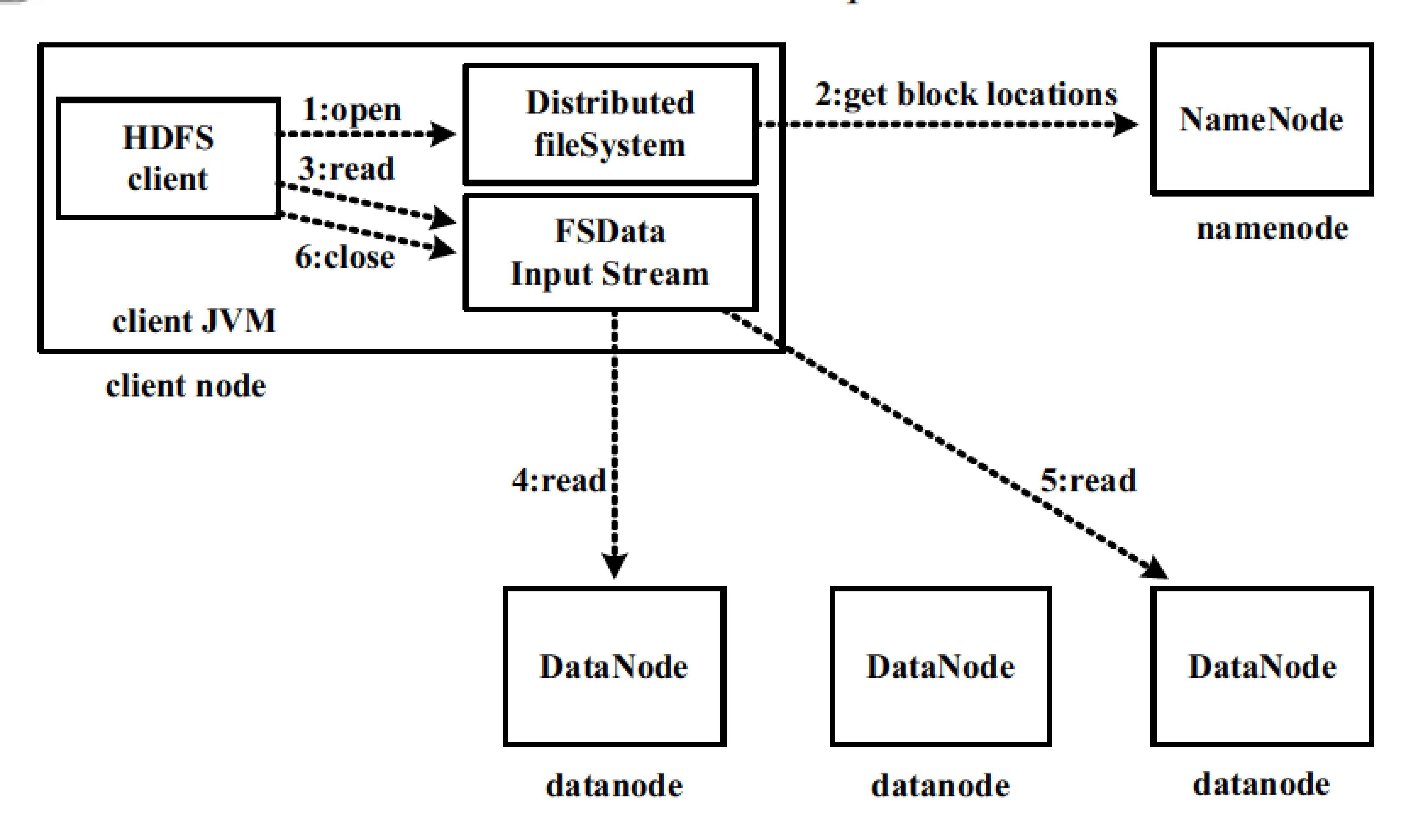


图 3.6　从 HDFS 读文件流程

（2）写文件流程

如图 3.7 所示，对 HDFS 进行写操作过程也相对比较复杂，需调用 client、NameNode、

DataNode 3 个进程以及操作相关的 DistributedfileSystem、FSDataOutputStream 等类。

步骤 01 client 调用 DistributedfileSystem 的 create 方法，创建文件。

步骤 02 DistributedfileSystem 通过 RPC 调用 NameNode，创建一个文件到文件系统的命名空间，并将 FSDataOutputStream 返回给 client。

步骤 03 client 向文件写数据，写数据块时，DFSOutputStream 把数据分成若干数据包（packet）；FSDataOutputStream 询问 NameNode，找到存储这个数据块以及副本的 DataNode 列表；该 DataNode 列表组成一个管道，由 3 个 DataNode 组成。

步骤 04 FSDataOutputStream 首先把数据包写入管道的第一个 DataNode，然后管道把数据包转发给第二个 DataNode，类似地依次转发到最后一个 DataNode。

步骤 05 当管道里所有 DataNode 都写入成功时，当前数据包的写操作完成，发送应答给 FSDataOutputStream，开始写下一个数据包。

步骤 06 所有数据块的写操作结束后，client 调用 FSDataOutputStream 的 close 方法，关闭该新建文件。

步骤 07 FSDataOutputStream 通知 NameNode，当前文件的写操作结束。

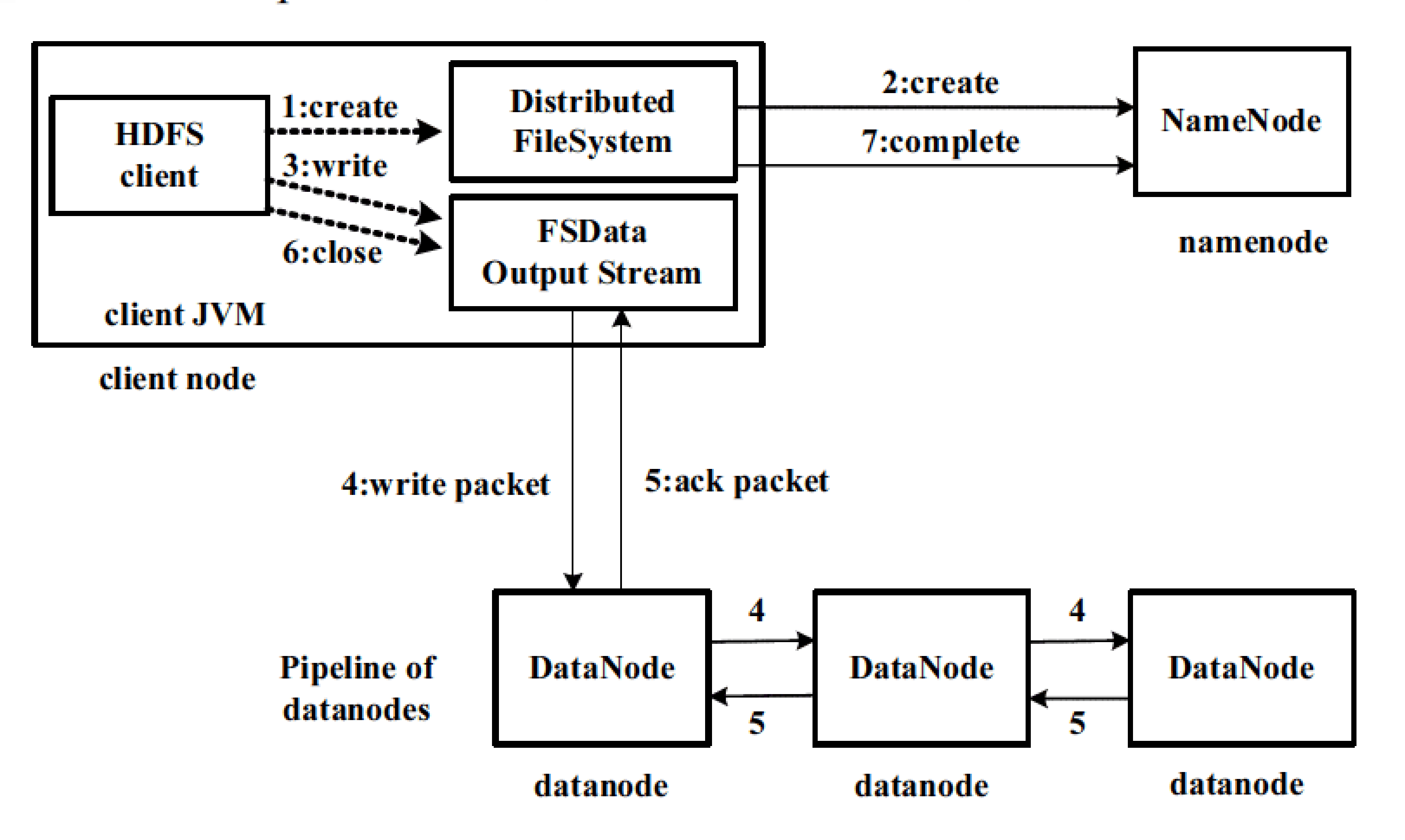


图 3.7　向 HDFS 写数据流程

3.3.2　MapReduce 模型

随着视频分辨率的提升、视频采集系统规模的扩大以及视频处理算法越来越复杂，海量视频管理系统对计算能力的要求越来越高，并且算法较难并行化。

分布式计算以大量普通计算机为基础组建集群，通过高效的计算架构在集群上分布和调度处理任务，进一步通过并行处理方式使计算机集群达到或超过单个大型计算机的

处理能力。集群由同构计算节点互联构成，通过设计调度和冗余处理策略来处理可能发生的单点故障问题，实现负载平衡，保障系统效率和运行稳定。

海量视频管理需要处理的数据量庞大，算法类型差别很大，分布式计算集群的高度可扩展性和单节点的通用性优势明显，通过增加或减少集群内的计算机数量，并加以简单配置，可灵活满足处理要求。

1. MapReduce 计算流程

Dean J 和 Ghemawat S 在 2004 年首次提出 MapReduce 分布式计算模型，当时的设计需求用于进行网站日志文件分析；在此基础上，Google 的 Hadoop 项目实现了该计算模型。

MapReduce 模型主要通过 Map 和 Reduce 函数实现数据处理的大规模并行化，其计算过程包括两个阶段。Map 函数将输入数据进行切分，并映射到不同键值之上，组成键值对，被发送给集群内的主机进行处理，生成的中间结果以新的键值对的形式保存在集群内。Reduce 函数则收集具有相同键值的中间结果，并进行综合，得到最终输出。

MapReduce 模型借鉴函数式程序设计语言的思想，把集群中的分布式并行运算抽象为两个阶段，即 Map 函数阶段和 Reduce 函数阶段，并将并行化、容错、数据分布等细节对使用者进行隐藏，其执行过程如图 3.8 所示。

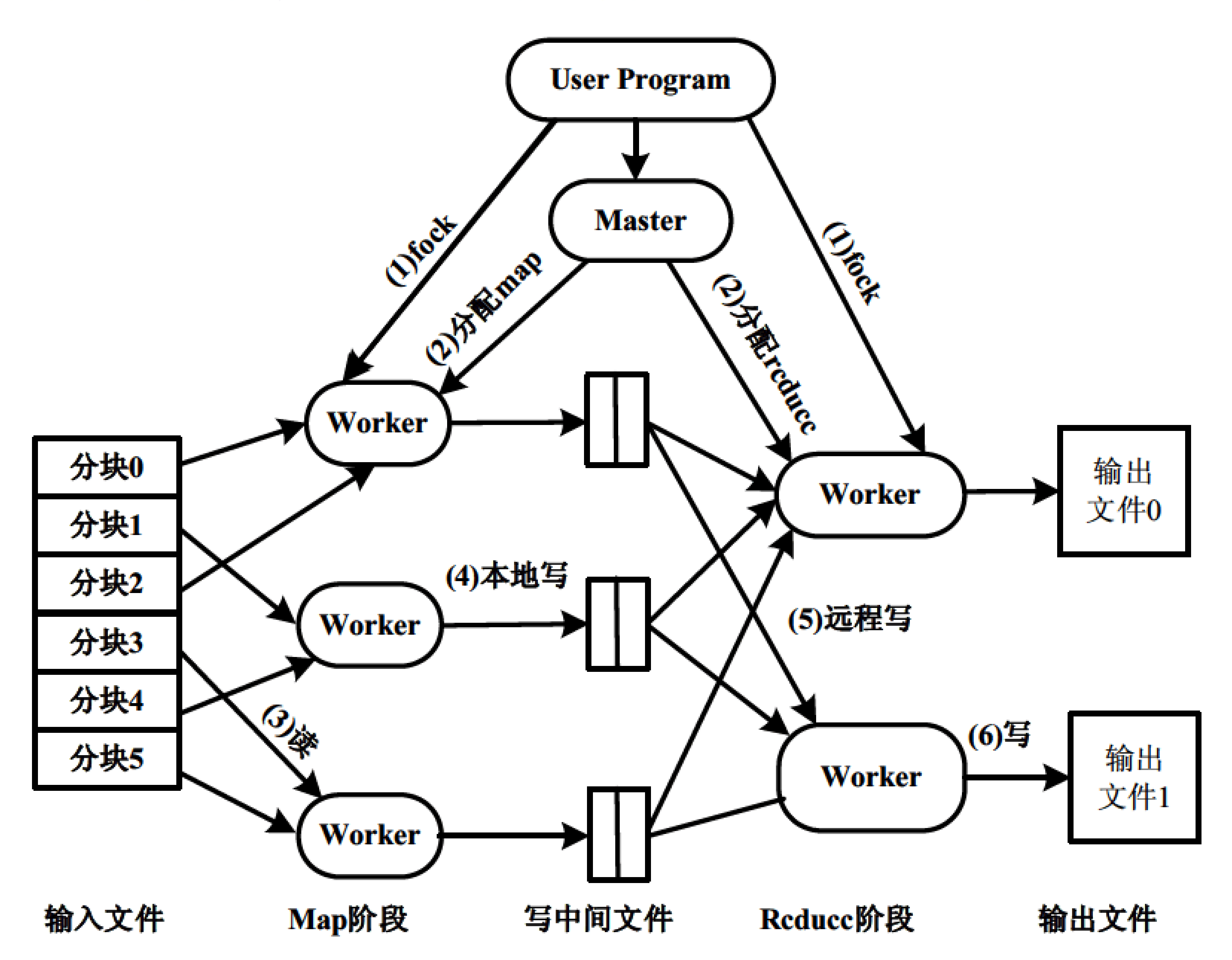


图 3.8 MapReduce 执行过程

MapReduce 执行 6 个步骤，可以简化为 3 个阶段。

- 输入阶段

Map-Reduce 库首先将输入文件分割成 M 片，每个片的大小在 16MB~64MB 之间；然后在集群中随机大量拷贝。

- Map 阶段

这些拷贝程序中的主节点（master）分配 map 任务和 reduce 任务，被分配 map 任务的工作节点读取输入片，从中解析出 key/value 对，由用户自定义的 Map 函数处理 key/value 对，产生中间 key/value 对。

- Reduce 阶段

Reduce 函数将传来的中间 key/value 对合并，并输出 R 个文件。

2. 视频并行处理模型

视频处理算法可以分为若干不同模块，通过对数据的管理和传递，共同完成处理任务，如视频图像增强、感兴趣区域提取、目标识别、目标跟踪等。这种模块称为算子，单个算子的输入数据可能是单帧图像或数帧图像，也可能是其他算子的输出数据。分布式集群上处理的一个任务由算子和其输入数据构成，利用算子编号和输入算子最后一帧的帧编号唯一确定该任务，标记为 Task(operatorID,frameID)。算子之间存在输入输出关系，表示算子之间存在连接。第 N 帧的 o1 算子可能需要前后几帧 o2 算子的输出作为输入，如通过帧差法提取运动目标，这种情况标记为：

$$\mathrm{Task}(\mathrm{o2},\mathrm{N}-\mathrm{d}):\mathrm{Task}(\mathrm{o2},\mathrm{N})\rightarrow\mathrm{Task}(\mathrm{o1},\mathrm{N})$$

$\mathrm{Task}(\mathrm{o2},\mathrm{N}-\mathrm{d})$到$\mathrm{Task}(\mathrm{o2},\mathrm{N})$称为 Task(O1,N)的输入任务，d 称为连接的深度，反映后续算子对前续算子历史信息的追溯深度。

算子之间的连接关系描述为：

$$\mathrm{o1}\rightarrow\mathrm{o2}\,,\quad \mathrm{depth}=\mathrm{d}$$

海量视频管理通常需要处理较长的历史信息，在这种情况下，d 趋向于无穷大。常用统计算子内置缓存用于记录历史信息，可以等效看做该算子向自身输出处理结果，被称为嵌套算子。任务和算子的依存关系如图 3.9 所示。

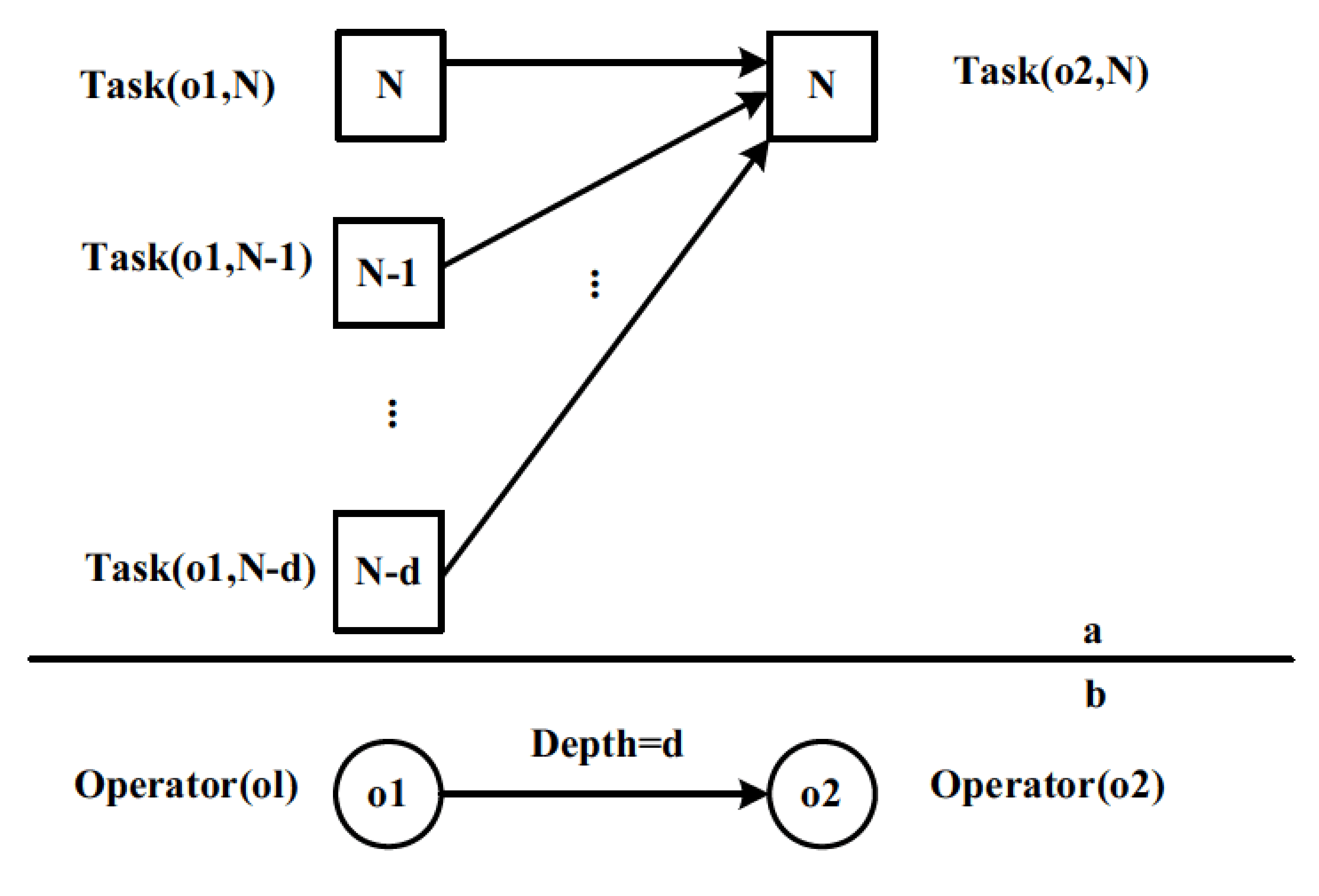


图 3.9 任务与算子关系图

视频处理算法可以由算子和算子之间的连接关系来表示，对于非嵌套算子 o1，第 N+1 帧的任务 Task(o1,N+1)不必等到 Task(o1,N)结束，而是只要其输入任务结束就可以和 Task(o1,N)并行处理，其帧率取决于其输入算子的帧率。如果整个视频算法全部由非嵌套算子构成，并且有足够多的并行主机分别用于计算，其处理速率取决于输入视频的帧率，可实现实时处理。从某一帧图像输入，经过多个算子串行连接处理将存在一定的延迟。任务并行处理的时序如图 3.10 所示。

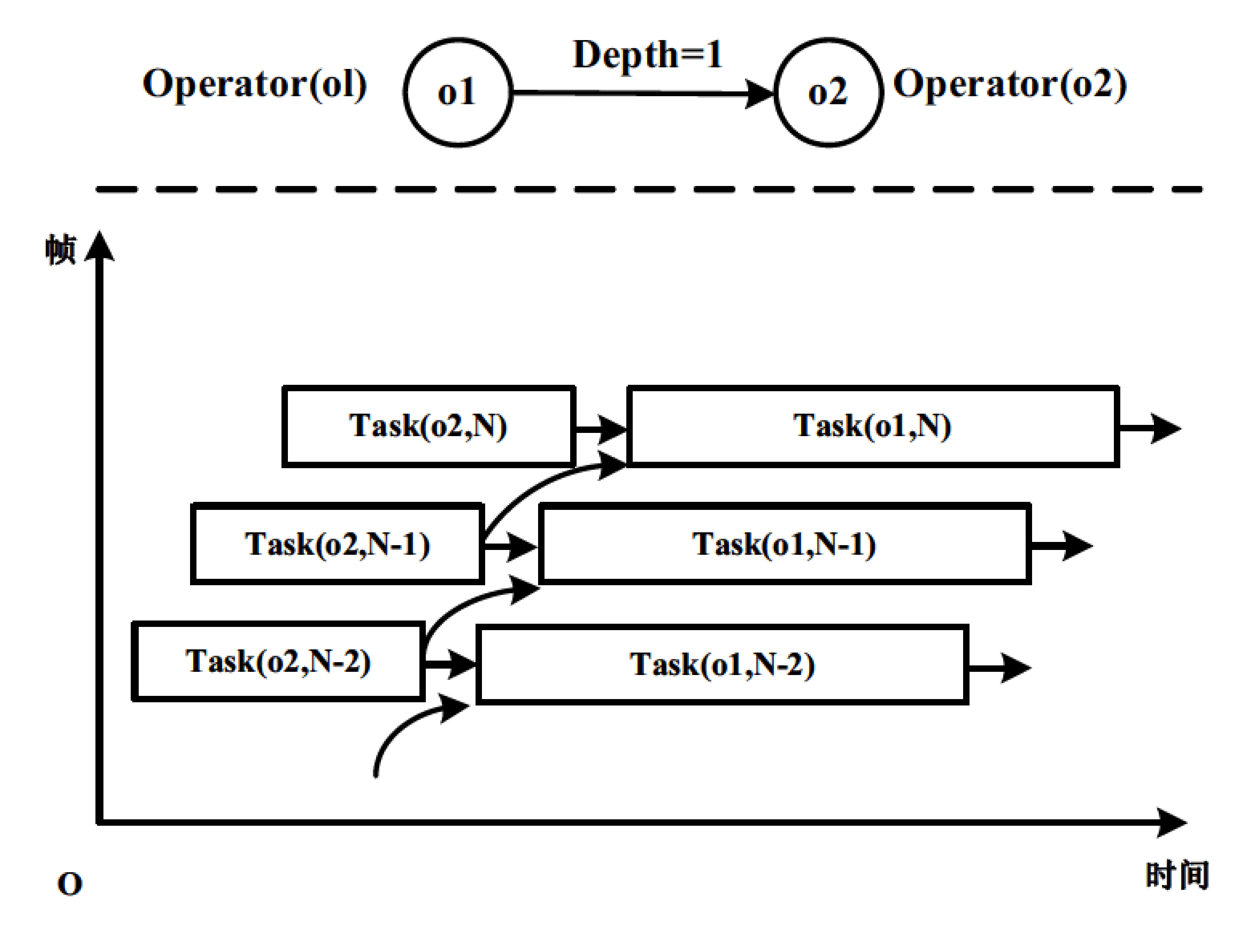


图 3.10 任务处理时序

对于嵌套算子，由于其需要自身的处理结果作为输入，分布式处理并不能提高该环节的效率，因此嵌套算子的单帧处理时间构成分布式视频处理模型的瓶颈，算法实现时需要优化算子结构，尽量减少嵌套算子。通常，可利用嵌套算子进行简单统计运算后作为整个系统最后的输出，计算量较小，对系统性能影响不大。

在分布式计算模型中，非嵌套算子相当于 MapReduce 模型的 Map 函数，算子号和帧号构成其键值，由集群内的主机并行处理。嵌套算子相当于 Reduce 函数，综合非嵌套算子的处理结果，生成系统最终输出。分布式计算集群的层次结构如图 3.11 所示。

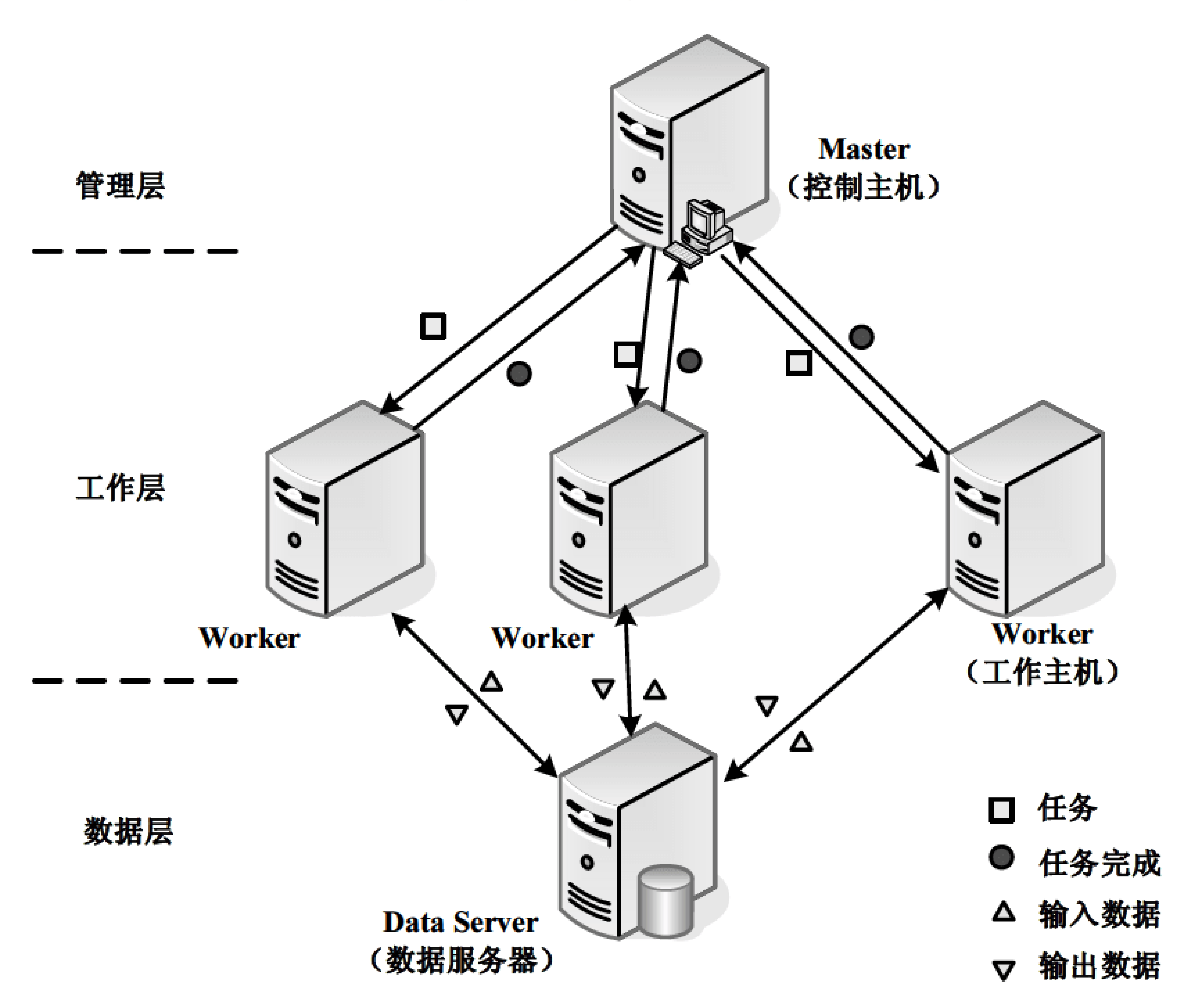


图 3.11　分布式计算群结构

管理层为 Master 主机，Master 分析算子之间的连接关系，合理生成计算任务，并按顺序发送给对应的工作主机处理，同时还要兼顾容错和负载均衡处理。工作层包含多台 Worker 主机，承担实际的计算任务。Worker 主机接收 Master 分配的计算任务，处理完成后将中间结果发给数据层。数据层主要由 DataServer 服务器构成，其作用是为工作层主机提供数据交换处理。中间数据以键值对的形式被存储在服务器端。

视频数据的分布式处理同样基于任务实现。任务对应其关联的算子号和帧号，该对应关系具有唯一性。 Worker 主机执行相关计算任务，将处理结果或中间数据发送到 DataServer 缓存，算法实际执行流程由算子间的连接关系决定。Worker 每完成一个任务，

便发送一个对应的通知消息给 Master。Master 检查该任务与其他任务的关系，并将尚未处理的后续任务加入排队列表，进一步地，Master 将等待列表中全部前置任务已经完成，即输入数据已经处理完毕的任务移至发送队列，等待时机发送给目的 Worker 处理，如图 3.12 所示。

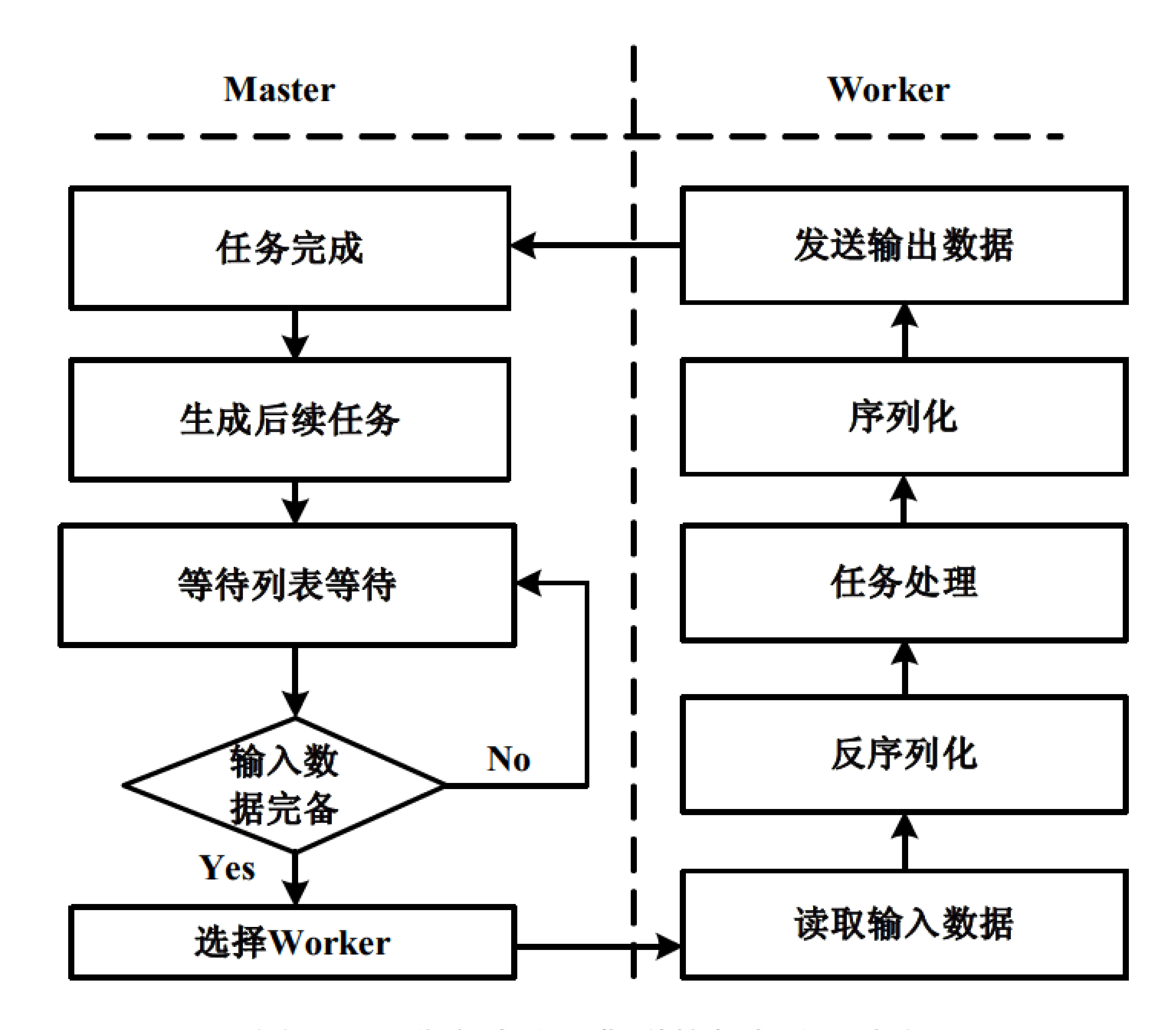


图 3.12　分布式处理集群的任务处理流程

基于 MapReduce 模型的分布式视频管理平台从嵌套算子和非嵌套算子的角度分解视频处理算法，属于时域分解，以帧为最小单位在集群内分配计算负载，实现分布式视频分析。最后，对于密集型任务，平台处理能力随集群内节点的计算能力线性增长；对现阶段常用的视频处理算法进行优化分割后，可基于该分布式视频处理平台实现实时处理。

3.4　博世视频管理系统

博世视频管理系统（Bosch Video Management System）是一款企业级视频安防解决方案，可以在任何 IP 网络之间提供无缝的数字视频、音频及数据管理。可以与相关视频监控产品配合使用，组成完整的视频安防管理系统，如图 3.13 所示。

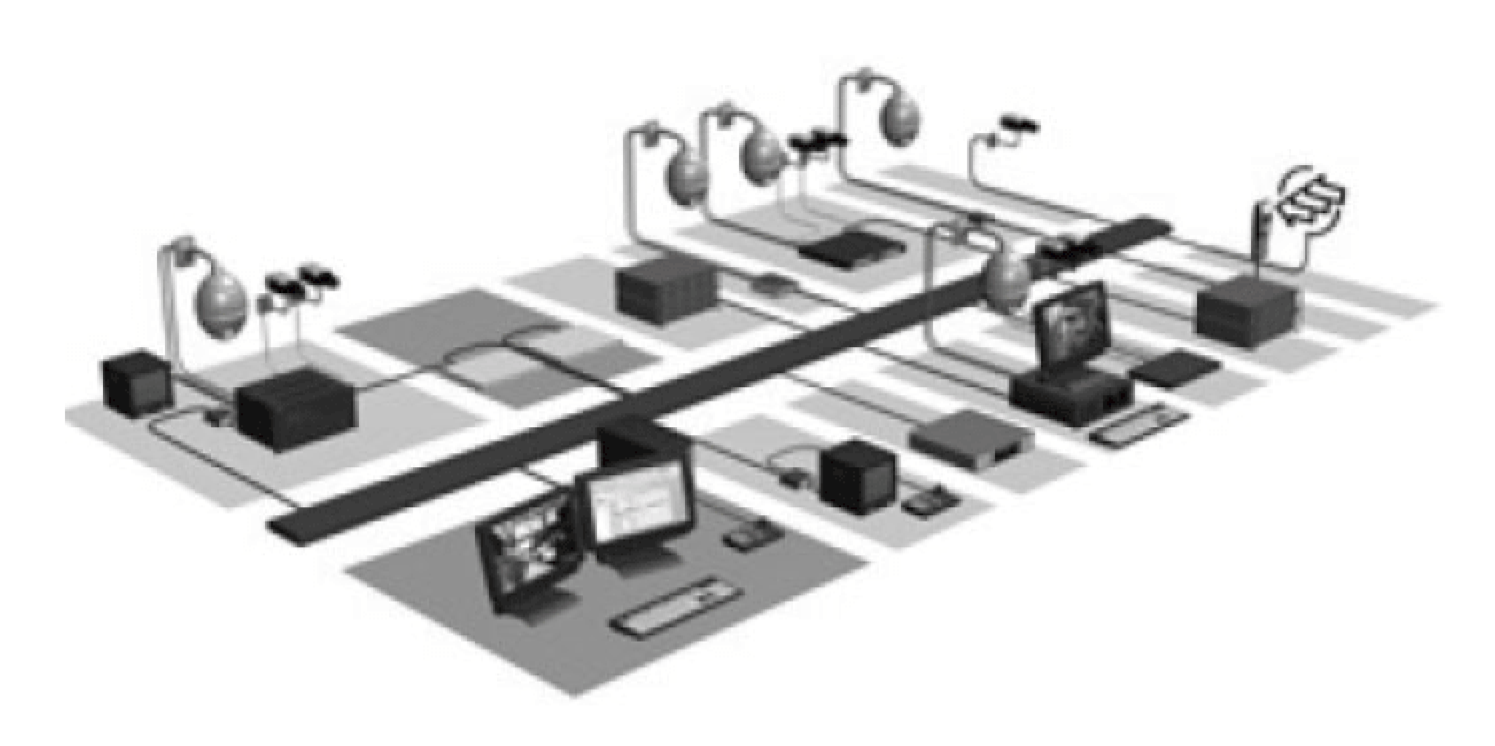


图 3.13　博世视频管理系统

1. 博世视频管理系统的功能

博世视频管理系统提供分布式网络视频解决方案，使用户无须安装专用的网络录像机（NVR），支持基于 iSCSI 技术的存储系统和 IP 视频设备，引入存储虚拟层的概念，可以像管理单个“虚拟”公用存储池一样，管理整个系统中的所有磁盘阵列，实现存储空间的智能分配。

用户无须安装相关的服务器硬件、操作系统、防病毒软件以及相关补丁，安装、操作和维护非常简单，降低了管理成本。

2. 博世视频管理系统的特点

博世视频管理系统具有如下特点：

- 基于客户端/服务器的企业级 IP 视频管理系统；
- 在系统范围内进行用户管理、报警处理、状态监视；
- 全面的虚拟矩阵功能，与原有模拟系统无缝融合；
- 借助报警优先级和可选的用户组分发功能处理报警；
- 通过先进的用户界面概念实现高效操作；
- 与标准计算机服务器、工作站和存储设备兼容。

3.5　微博视频管理系统

微博视频管理系统用于在互联网上检索与某段视频相似的视频在微博上的传播情况，支持视频的模糊搜索功能，在视频进行格式、帧率、清晰度、切分组合、LOGO 添加等变换后仍能准确地进行检索。

微博视频管理系统基于先进的分布式云计算架构，结合高性能的视频采集、解码、

图文提取与比对技术，实现对微博中传播视频的快速检索和分析，其结构如图 3.14 所示。

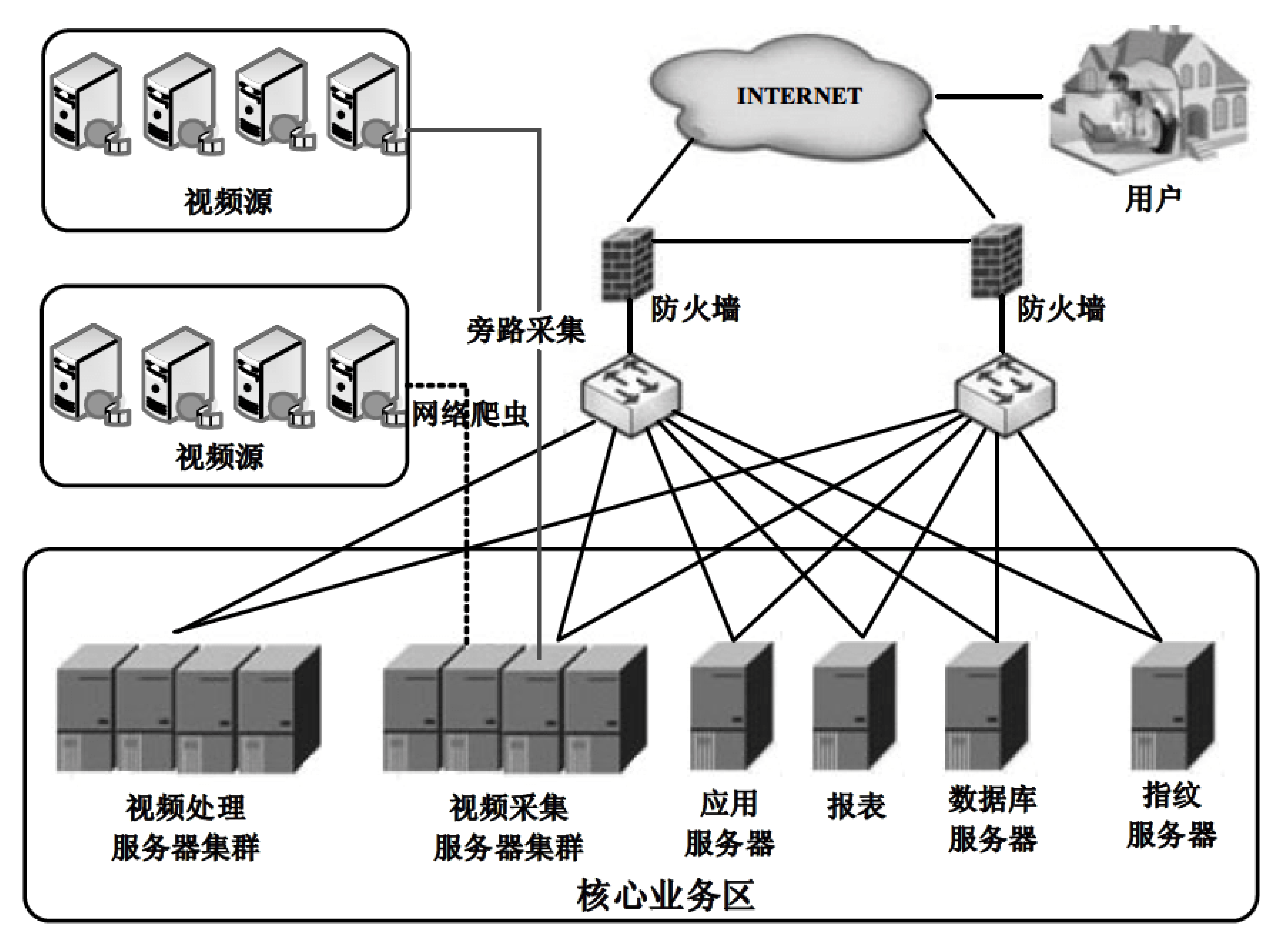


图 3.14　微博视频管理系统

1. 微博视频管理系统的功能

- 视频内容监测

用户提交原始视频，系统检索出微博中与此视频相关的全部视频。

- 热点事件追踪

把一批视频定义为一个事件集合，系统检索该事件的所有相关视频，分析该事件的传播轨迹、事件趋势、影响范围。

- 视频聚合

通过语义分析和图文对比相结合，实现视频内容的自动聚合与分类。

- 重点账号监控

对微博的重点账号进行全程监控，对重点人物的活跃度、倾向性、威胁度进行全方位分析和监控。

2. 微博视频管理系统的特点

- 海量视频处理
  - 基于 Hadoop 的云计算架构；

- 支持 PB 级视频数据的存储与分析。

- 视频分析引擎
  - 基于人眼视觉感知模型的视频图像处理技术;
  - 抗干扰能力强，在视频进行各类变化后仍能有效识别;
  - 单台视频处理引擎大于 200GB/h;
  - 视频搜索时间小于 5 秒。

3.6　VOD 视频点播及管理系统

VOD 系统通过网络向分布在各处的终端设备实时、定时触发等多种形式，发布滚动通知和视频信息等内容。如图 3.15 所示，该系统由流媒体服务器、管理服务器、播放端组成，采用 B/S 管理模式，实现广域网、局域网的整合控制；可支持多种终端接入，支持高清节目播放，应用各种封装技术提供完善接口，进行方便、快捷的定制开发。

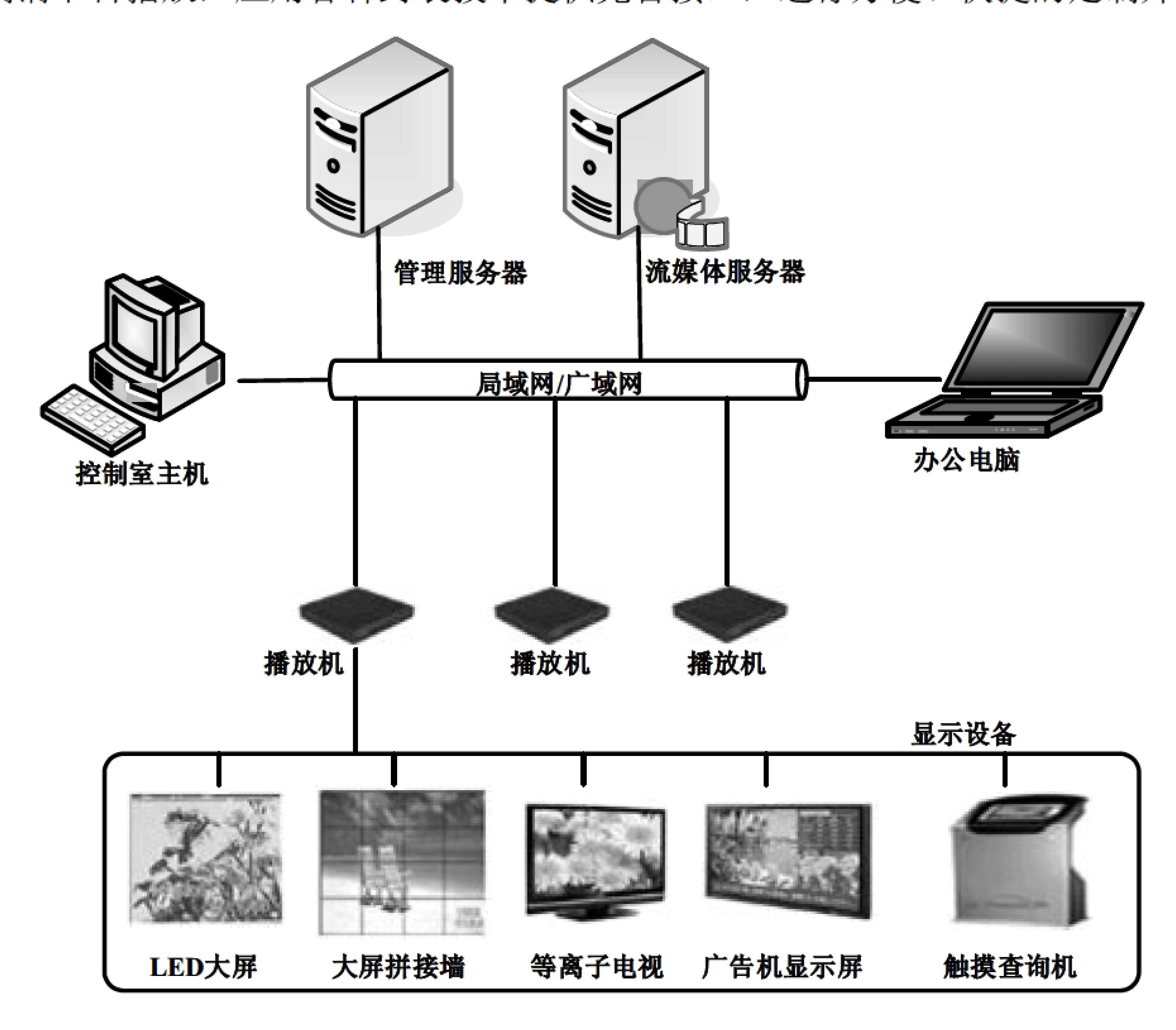


图 3.15　VOD 系统

1. VOD 系统的功能

- 业务功能

支持高清视频点播，提供良好的视听感受；采用导演、演员、发行年代、类型等多种筛选条件，为用户查询节目时提供方便、简单和快捷的操作方式。

- 管理功能

支持宣传资料管理、节目管理、信息管理、配置管理、系统设置、播放监控、操作日志查询等功能。

2. VOD 系统的特点

- 开放的业务平台

采用开放理念提供标准的外部接口，支持内容、信息、服务等提供商接入，支持广告、信息等平台接入，为系统建立统一资源平台。

- 分级管理

分析产业链中各个角色的需求，全面体现角色管理功能。

- 灵活的部署方案

支持集中式、分布式、混合式部署方案，满足各种规模需求。

- 采用 Linux 操作系统

Linux 系统具有极高的安全性和稳定性。

第 4 章

海量视频分析

海量视频分析是面向海量视频的深度应用，采用计算机视觉、视频图像处理、人工智能、机器学习、应用数学等学科的理论和方法，对海量视频进行格式解析、特征提取、数据管理、快速分类等。

本章针对海量视频分析需求，重点介绍常用理论和基本方法，包括 Harris 描述子、SIFT 描述子、K-Means 方法、K 近邻法、SVM 方法、BP 网络、多感知器模型、CNN、AdaBoost 方法、模拟退火和遗传方法。

4.1 Harris 描述子

角点特征能够减少用于计算的数据量，同时不损失用于描述主要特征的其他重要信息。学者 Chris Harris 于 1988 年提出了著名的 Harris 算子，该算子是一种有效的角点特征提取方法。

1. 基本原理

Harris 算子继承 Moravec 算子的思想精髓，并做出重要改进。Harris 算子可从连续角度进行推导，考虑每个方向上的自相关性，使用圆形模板窗代替 Moravec 算子的方形窗。

Harris 算子取以目标像素点为中心的一个小窗口，计算窗口沿任何方向移动后的灰

度变化，并用解析形式表达。

设以像素点(x, y)为中心的小窗口在 x 方向上移动 u，y 方向上移动 v，Harris 给出灰度变化度量的解析表达式：

$$E(u,v)=\sum_{x,y}w(x,y)[I(x+u,y+v)-I(x,y)]^2$$

其中，$w(x,y)$为窗口函数。

如图 4.1 所示，Harris 算子采用高斯函数作为窗口函数，离中心点越近的像素具有越大的权重，从而可以减少噪声影响。

$$w(x,y)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}$$

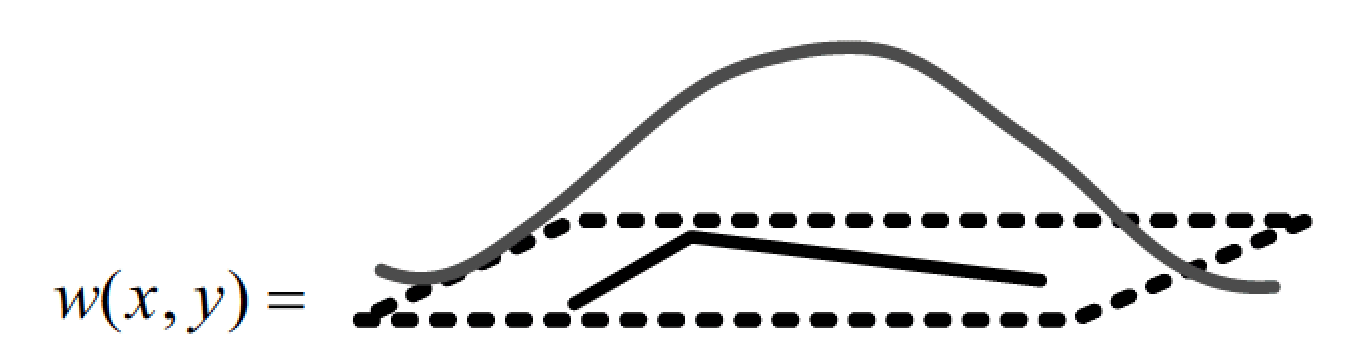


图 4.1　高斯函数

Moravec 算子只考虑每隔 45° 方向，Harris 算子则通过 Taylor 级数展开可以逼近任意方向。

设 I 为图像灰度函数，I_x 为 x 方向的差分，I_y 为 y 方向的差分。Taylor 展开为：

$$I(x+u,y+v)=I(x,y)+I_x u+I_y v+O(u^2,v^2)$$

因此可得：

$$\begin{aligned}E(u,v)&=\sum_{x,y}w(x,y)\left[I(x+u,y+v)-I(x,y)\right]^2\\&=\sum_{x,y}w(x,y)\left[I_x u+I_y v+O\left(u^2,v^2\right)\right]^2\end{aligned}$$

略去二次以上高阶无穷小项，有：

$$E_{x,y}=\sum w_{x,y}[u^2(I_x)^2+v^2(I_y)^2+2uvI_xI_y]=Au^2+2Cuv+Bv^2$$

将 $E_{x,y}$ 转化为二次型，有：

$$E_{x,y}=[u\ v]M\begin{bmatrix}u\\v\end{bmatrix}$$

M 为实对称矩阵：

$$M=\sum w_{x,y}\begin{bmatrix} I_x^2 & I_x \bullet I_y \\ I_x \bullet I_y & I_y^2 \end{bmatrix}$$

通过对角化处理得到：

$$E_{x,y}=R^{-1}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}R$$

其中，R 为旋转因子，对角化处理后并不改变以 u、v 为坐标参数的空间曲面的形状，其特征值反映了两个主轴方向的图像表面曲率。

矩阵 M 的两个特征向量 λ_1 和 λ_2，与矩阵 M 的主曲率成正比。如图 4.2 所示，Harris 算子利用 λ_1 和 λ_2 表征变化最快和最慢的两个方向，若两个特征向量都很大，则对应角点；若 λ_1 和 λ_2 一大一小则对应边缘；若 λ_1 和 λ_2 都小则对应变化缓慢的平坦区域。

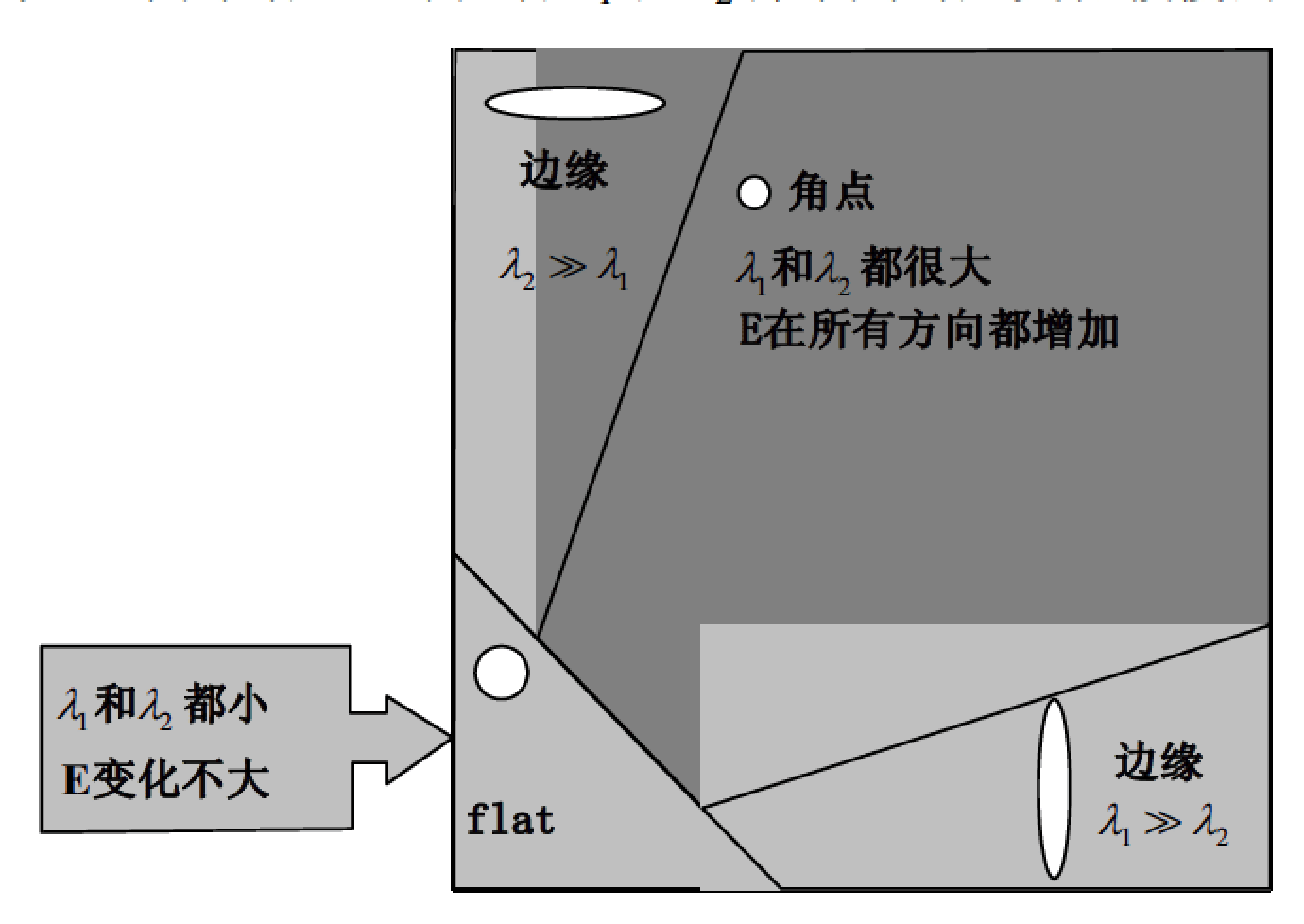


图 4.2　用矩阵 M 的特征向量分类图像像素点

求解特征向量的计算量比较多。由于两个特征值的和对应于矩阵 M 的迹，它们的积等于矩阵 M 的行列式，所以常用如下的角点响应函数（*CRF*）来判定角点。

$$R=\det M-k\left(\operatorname{trace} M\right)^2$$

其中，det M 表示 M 的行列式，trace M 表示 M 的迹，k 常取 0.04~0.06。当目标像素点的 *CRF* 值大于给定阈值时，该像素点即为角点。

2. 实现方法

Harris 算子求取的流程如下。

步骤01 对每个像素点计算相关矩阵 M；

$$A = w(x,y) \otimes I_x^2$$

$$B = w(x,y) \otimes I_y^2$$

$$C = D = w(x,y) \otimes (I_x I_y)$$

$$M = \begin{pmatrix} A & D \\ C & B \end{pmatrix}$$

步骤02 计算每个像素点的 Harris 角点响应函数；

$$R = (AB - CD)^2 - k(A+B)^2$$

步骤03 在 $w \times w$ 范围内寻找极大值点，若 Harris 角点响应大于阈值，则视为角点。

3. 算法特点

Harris 算子计算简单，只用到灰度的一阶差分以及滤波。Harris 算子对图像中的每个点都计算其兴趣值，然后结合邻域选择最优点。在纹理信息丰富的区域，Harris 算子可以提取出大量有用的特征点；反之则特征点较少。

如图 4.3 所示，Harris 算子对图像平移、旋转、灰度变换、噪声干扰和视角变化有较强的适应性，是一种比较稳定的点特征提取算子。

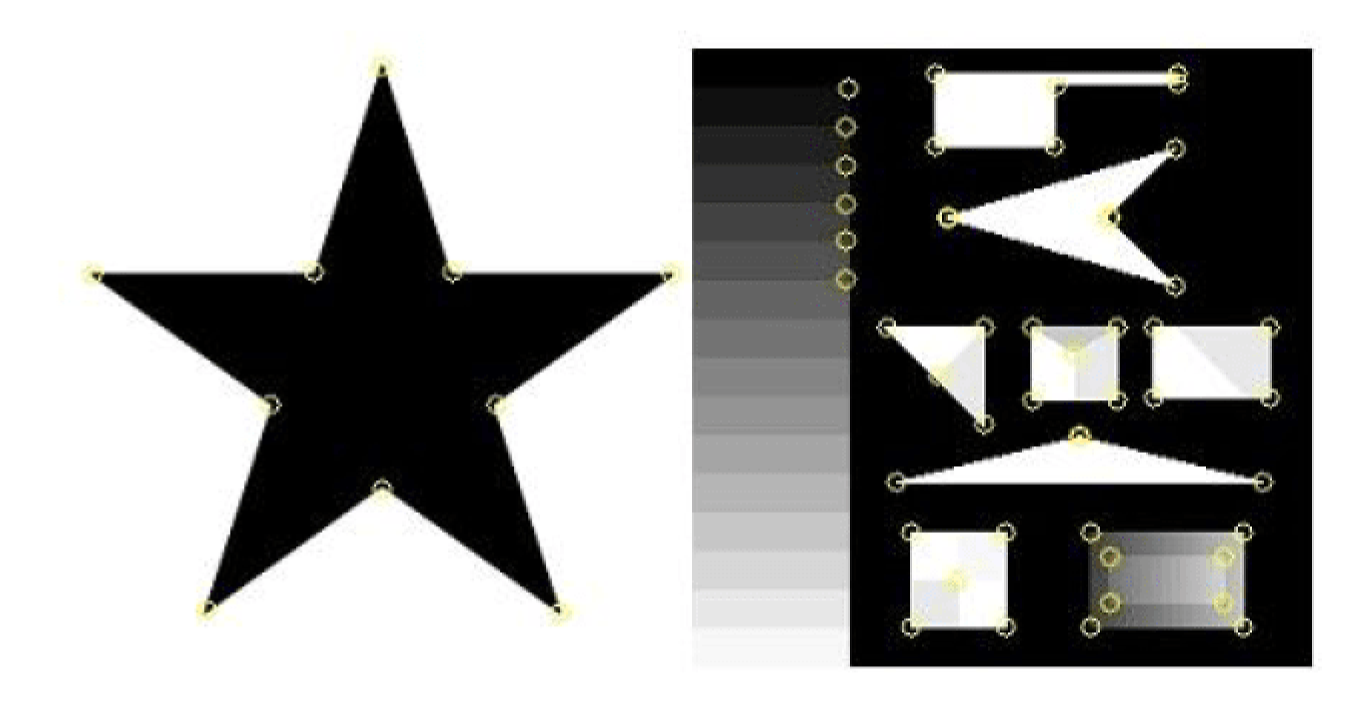


图 4.3 Harris 算子对简单图像的响应

如图 4.4 所示，Harris 算子对尺度很敏感，在某个尺度下是角点，在另一个尺度下可能就不是。

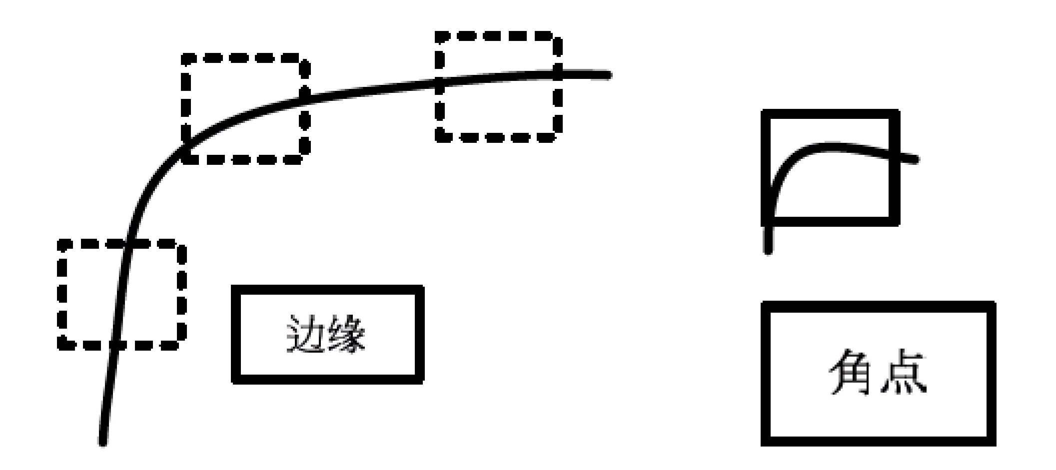


图 4.4　Harris 算子对尺度的敏感性

4.2　SIFT 描述子

SIFT 描述子（Scale Invariant Feature Transform，尺度不变特征变换）由加拿大哥伦比亚大学（University of British Columbia，Canada）的 David 在 1999 年的 ICCV 会议上提出。它是一种基于尺度空间的，对图像平移、旋转、缩放以及一定视角和光照变化等保持不变性的局部特征描述子。

1. 尺度空间的生成

尺度空间理论模拟图像数据的多尺度特征，为了使特征具有尺度不变性，特征点的检测在多尺度空间完成。

高斯卷积核是实现尺度变换的常用变换核，并且是唯一的线性核。二维图像的尺度空间定义为：

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y)$$

其中，$G(x,y,\sigma)$ 是尺度可变高斯函数：

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)}/2\sigma^2$$

其中，符号“*”表示卷积，(x, y)代表图像的像素坐标，σ是尺度空间因子，σ越小表示图像平滑越少，相应尺度越小。大尺度对应于图像的整体特征，小尺度对应于图像的局部细节特征。

如图 4.5 所示，两组高斯尺度空间图像表示金字塔的构建过程，其中第二组图像通过对第一组图像进行下采样得到。

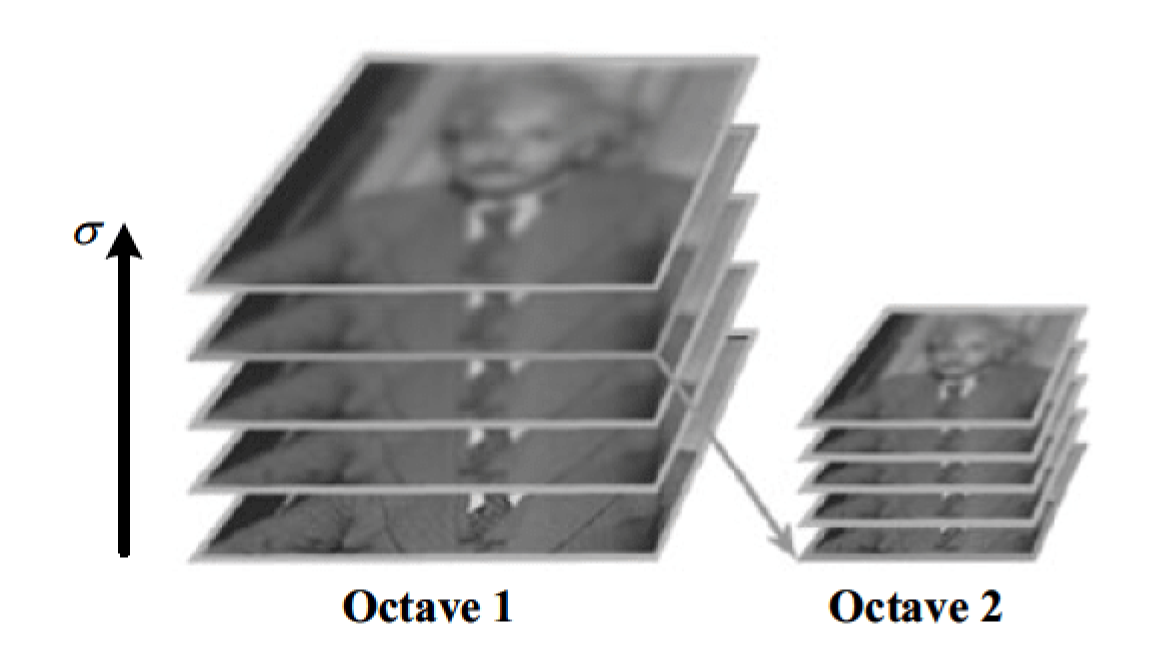


图 4.5　金字塔的构建示例

基于高斯差分尺度空间（DoG），利用不同尺度的高斯差分核与图像卷积处理，可以有效地在尺度空间进行稳定的关键点检测。

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned}$$

实际处理中用差分近似代替微分：

$$\sigma\nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma}$$

则有：

$$G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2\nabla^2 G$$

其中，k-1 是常数，不影响极值点位置的求取。

Lindeberg 与 1994 年发现高斯差分算子（Difference of Gaussian，DoG）与尺度归一化的高斯拉普拉斯算子 $\sigma^2\nabla^2 G$ 非常相似，如图 4.6 所示，图中实线表示高斯差分算子（DoG），虚线表示高斯拉普拉斯算子（LoG）。而高斯差分函数计算简单，效率高，可作为尺度归一化的高斯拉普拉斯算子（Laplacian of Gassian，LoG）的一种近似表示。

2002 年，Mikolajczyk 指出，与梯度、Hessian 或 Harris 特征提取函数比较，$\sigma^2\nabla^2 G$ 的极大值和极小值能够产生最稳定的图像特征。

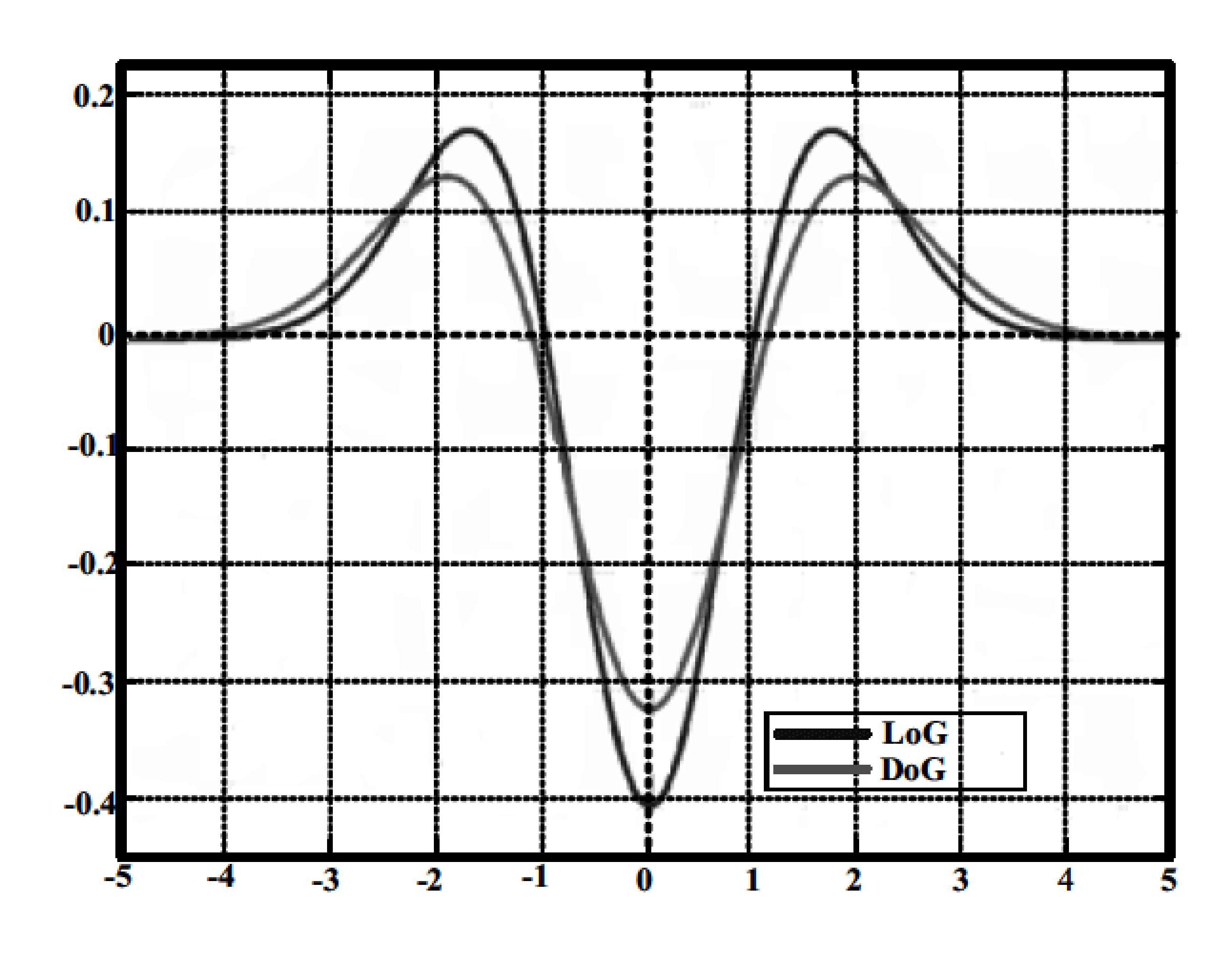


图 4.6　LoG 和 DoG 的比较

图 4.7 展示了构造 $D(x,y,\sigma)$ 的一种有效方法，具体介绍如下：

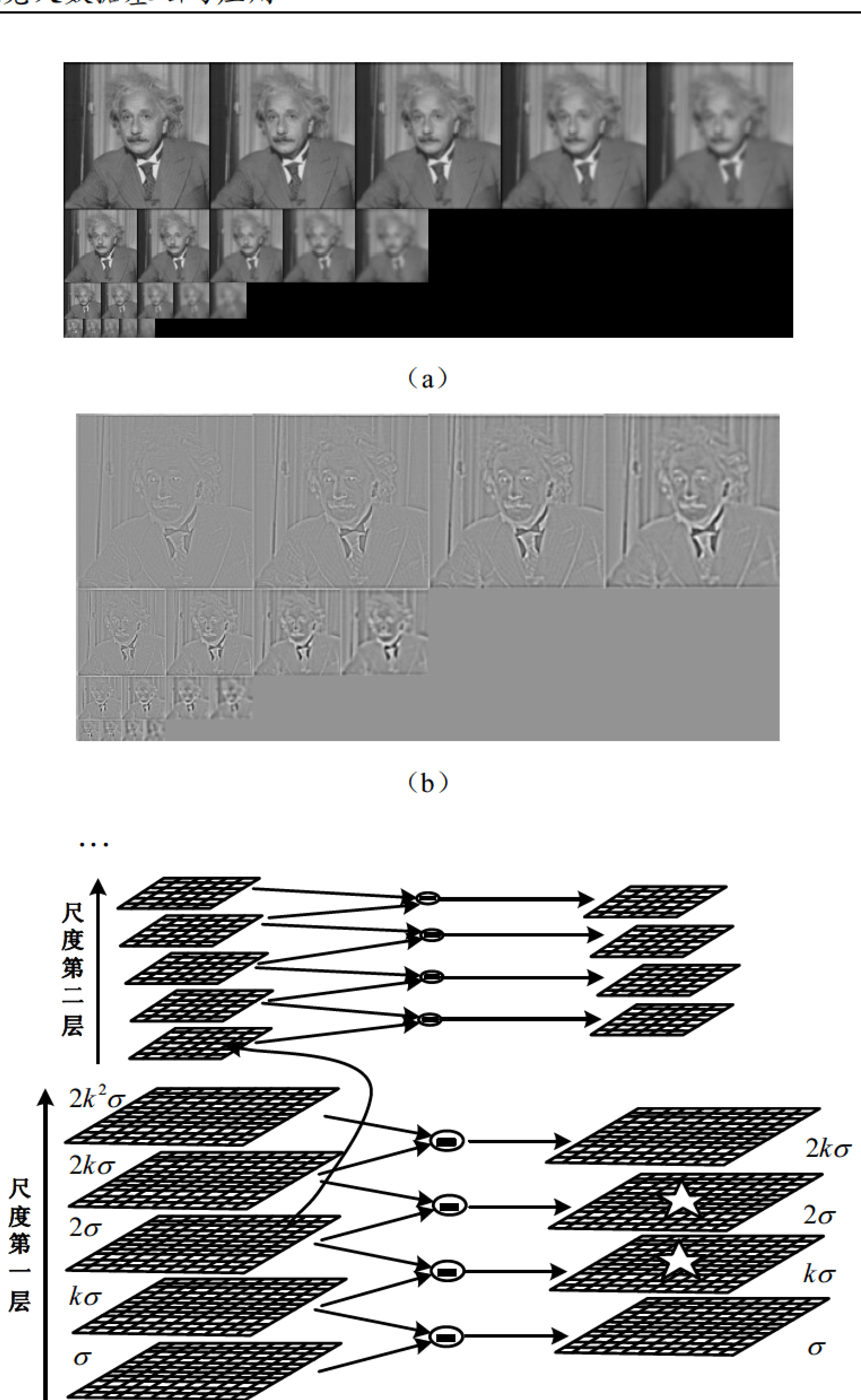
步骤 01　采用不同尺度的高斯核对图像进行卷积，得到图像的不同尺度空间，将该组图像作为金字塔图像的第一层。

步骤 02　对第一层图像的 2 倍尺度图像以 2 倍像素距离进行下采样，得到金字塔图像的第二层的第一幅图像，进一步地，用不同尺度因子的高斯核对该图像进行卷积，获得金字塔图像中第二层的一组图像。

步骤 03　与步骤 02 类似，以金字塔图像中第二层中的 2 倍尺度图像再次以 2 倍像素进行下采样，得到金字塔图像第三层的第一幅图像，采用不同尺度因子的高斯核对该图像进行卷积，获得金字塔图像第三层的一组图像。

以此类推，获得金字塔图像每一层的一组图像，如图 4.7(a)所示。将 4.7(a)每一层的相邻高斯图像相减，得到高斯差分图像，如图 4.7(b)所示。图 4.7(c)中右列显示每组的高斯差分图像。

最后，以高斯差分函数近似实现归一化的高斯-拉普拉斯函数，从高斯差分金字塔分层结构中提取图像极值点作为候选特征点。将 DoG 尺度空间中每个点与相邻尺度和位置的点逐一比较，检测到的局部极值位置即为特征点所处位置和对应尺度。



（a）

（b）

（c）

图 4.7　高斯金字塔中相邻尺度的两幅高斯图像相减得到 DoG 图像

2. 空间极值点检测

针对每一采样点，均要比较其与所有邻域点的差别，判断其是否为尺度空间的极值

点。如图 4.8 所示，将标注为“×”的检测点和它同尺度的 8 邻域点、上下相邻尺度分别对应的 18 个点共 26 个点比较，以确保检测到在原始图像空间和变尺度空间都满足条件的极值点。

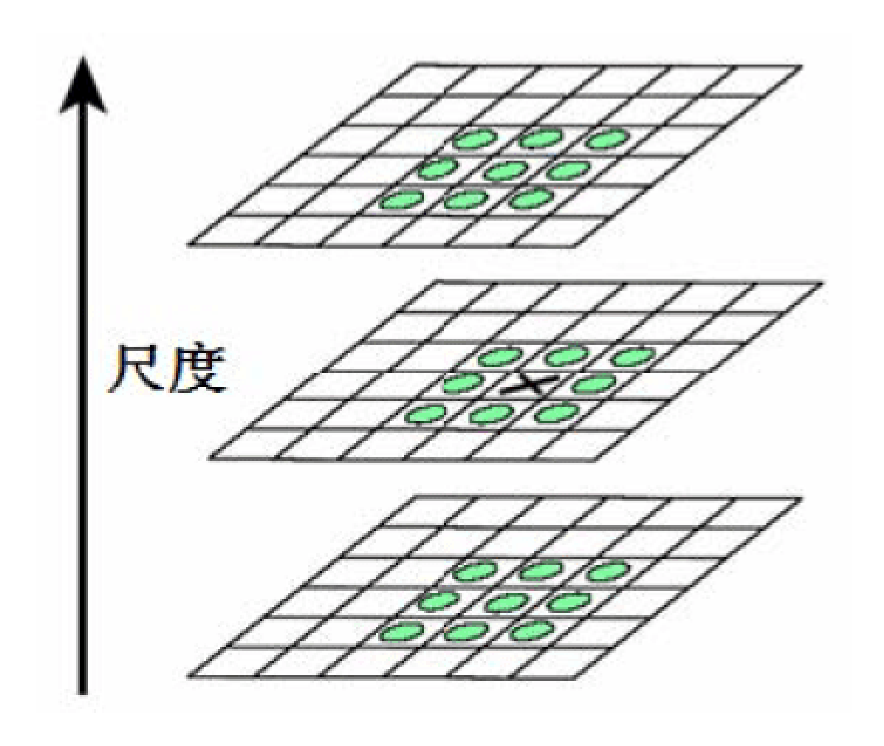


图 4.8　DoG 尺度空间中局部极值检测

因为需要同相邻尺度进行比较，在一组高斯差分图像中，只能检测到两个尺度的极值点，而其他尺度的极值点则在图像金字塔中上一层高斯差分图像进行检测。以此类推，最终实现在图像金字塔中不同层高斯差分图像中进行不同尺度的极值点检测。

因为某些极值点响应较弱，而且 DoG 算子会产生较强的边缘响应，上述处理得到的极值点并不都是稳定的特征点。

为考察点的特征性，可以获取对应点处的 Hessian 矩阵，主曲率通过一个 2×2 的 Hessian 矩阵 $\boldsymbol{H}$ 求出：

$$H=\begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

$\boldsymbol{H}$ 的特征值α和β代表 x 和 y 方向的梯度。

$$Tr(H)=D_{xx}+D_{yy}=\alpha+\beta$$

$$Det(H)=D_{xx}D_{yy}-(D_{xy})^2=\alpha\beta$$

其中，$Tr(H)$表示矩阵 $\boldsymbol{H}$ 对角线元素之和，$Det(H)$表示矩阵 $\boldsymbol{H}$ 的行列式。假设α是较大的特征值，而β是较小的特征值，令$\alpha=r\beta$，则

$$\frac{Tr(H)^2}{Det(H)}=\frac{(\alpha+\beta)^2}{\alpha\beta}=\frac{(r\beta+\beta)^2}{r\beta^2}=\frac{(r+1)^2}{r}$$

在实际处理中，可由采样点的相邻差估计导数。

$\boldsymbol{D}$ 的主曲率和 $\boldsymbol{H}$ 的特征值成正比，假设α为最大特征值，β为最小特征值，则当两

个特征值相等时$(r+1)^2/r$的值最小，随着r的增大而增大。$(r+1)^2/r$越大，说明两个特征值的比值越大，即在某一个方向的梯度值越大，而在另一个方向的梯度值越小，即边缘的情况。为了剔除边缘响应点，需要让该比值小于一定的阈值。为了检测主曲率是否小于某阈值 r，只需检测：

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$

一般取 $r=10$。

3. 关键点方向分配

为保证描述子的旋转不变性，需要依据图像的局部特征给每个关键点分配方向。可以计算梯度模值和方向如下：

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2}$$
$$\theta(x,y)=\tan^{-1}((L(x,y+1)-L(x,y-1))/L(x+1,y)-L(x-1,y)))$$

其中，L 为每个关键点各自的尺度。

完成上述计算后，在关键点的邻域窗口内采样，并统计直方图以获取邻域像素的梯度方向。梯度直方图的范围为 0°~360°，其中每 10° 为一个方向，共 36 个方向。

计算方向直方图时，需用参数σ对方向直方图进行加权，其中σ的取值为关键点所在尺度的 1.5 倍高斯权重。如图 4.9 中的圆形所示，中心处的颜色较浓，表示权值最大；边缘处颜色浅，表示权值较小。图中给出了 8 个方向的方向直方图示例。

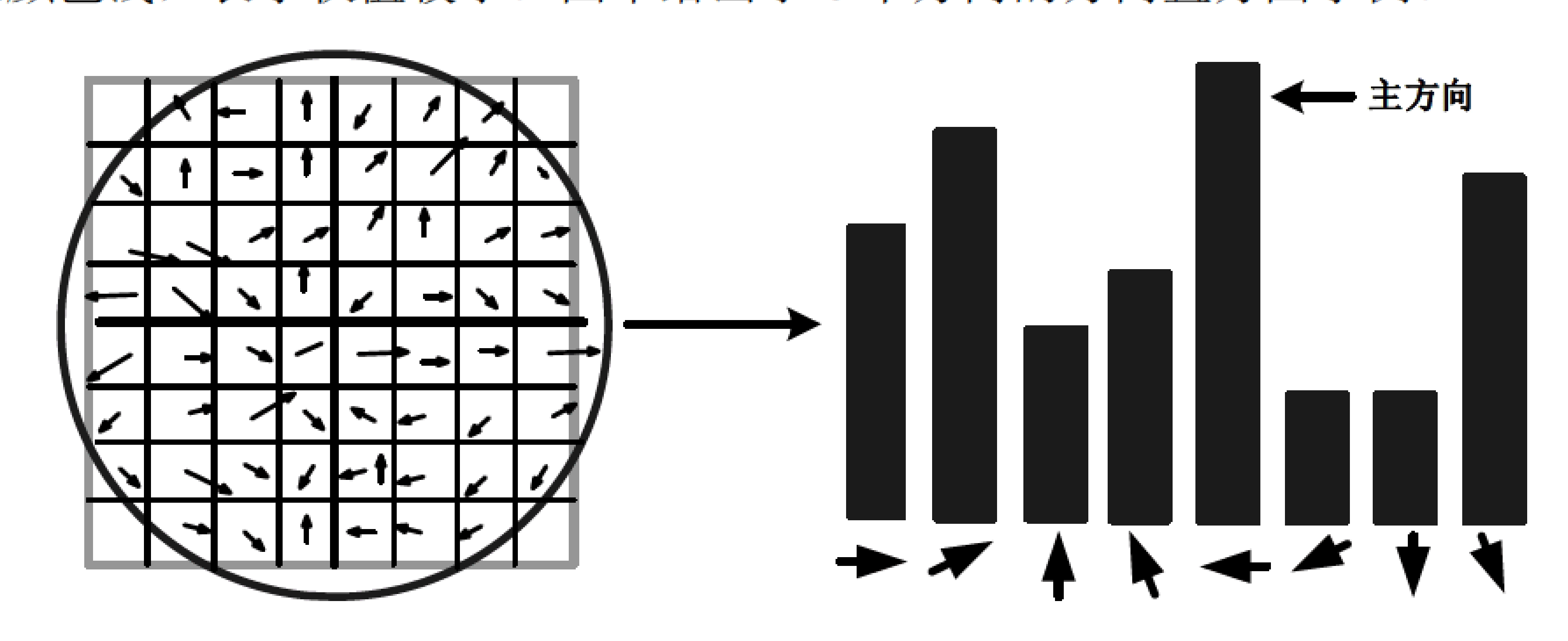


图 4.9　关键点方向分配

关键点方向可选用方向直方图的峰值，即该关键点邻域梯度的主方向。为增强匹配的鲁棒性，该关键点选择峰值大于主方向峰值 80%的方向作为辅方向。对于同一梯度值

具有多个峰值的关键点，在相同位置和尺度将会创建具有不同方向的多个关键点。为提高关键点匹配的稳定性，通常仅有 15%的关键点被赋予多个方向。

4. 特征点描述子的生成

每个关键点包含 3 类特性：位置、尺度、方向。针对每个关键点，应对其建立可表达上述特性的描述子，而且该描述子应具有鲁棒性，不随光照、视角等外界环境的改变而变化。此外，该描述子还应有较高的独特性，以提高特征点正确的概率。

首先，旋转坐标轴，使其与关键点方向一致，以确保旋转不变性。

其次，以关键点为中心取 8×8 的窗口。图 4.10 左部分的中央黑点为当前关键点的位置，圆圈表示高斯加权的范围，像素越靠近关键点，其梯度方向信息的作用越大。每个小格代表关键点邻域所在尺度空间的一个像素，箭头方向代表该像素的梯度方向，箭头长度代表梯度大小。

然后，在每个 4×4 的小块上计算 8 个方向的梯度方向直方图，描绘每个梯度方向的累加值，形成一个种子点，如图 4.10 右图所示。图中每个关键点由 4 个种子点组成，每个种子点有 8 个方向的信息。基于邻域方向信息联合可以增强算法抗噪声能力，同时对于含有定位误差的特征匹配具有较好的容错性。

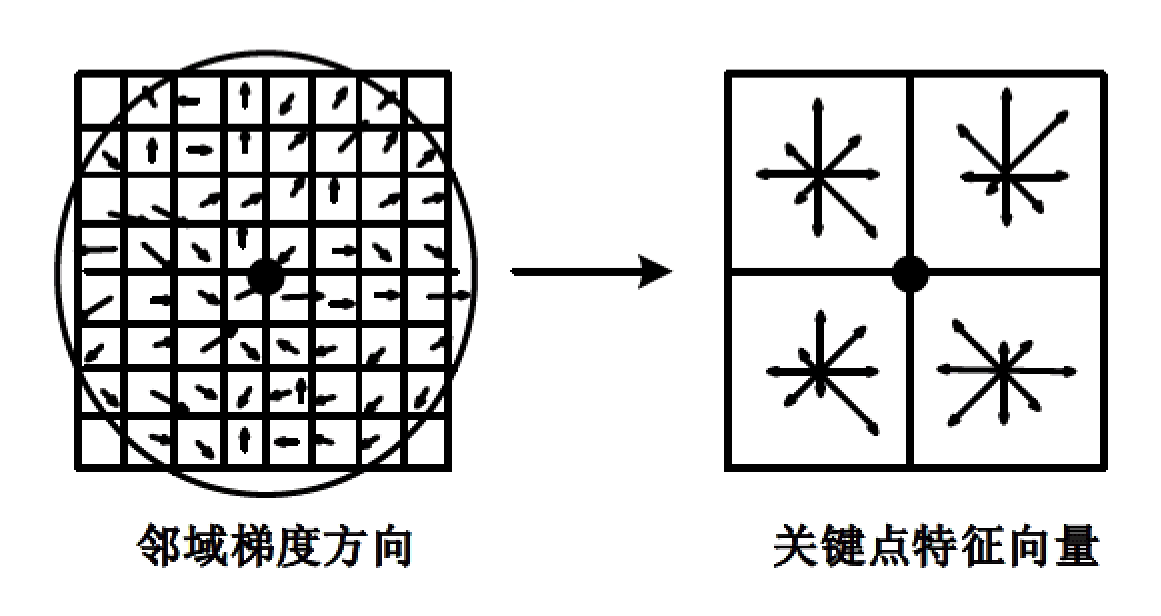


图 4.10　由关键点邻域梯度信息生成特征向量

最后，为了增强关键点匹配的鲁棒性，通常对每个关键点使用 4×4 共 16 个种子点来描述，每一个关键点可生成 128 维 SIFT 特征向量。通过上述处理提取到的 SIFT 特征向量已经去除尺度、旋转等几何变形因素的影响，再继续将特征向量的长度归一化，可进一步去除光照变化的影响。

5. 特征点匹配

将关键点特征向量的欧式距离作为两幅图像中关键点的相似度判定依据。取图像 1 中的某个关键点，并找出图像 2 中与其欧式距离最近的前两个关键点，如果最近的距离除以次近的距离小于某个比例阈值，则接受这一对匹配点。降低这个比例阈值，SIFT

匹配点数目会减少，但剩余匹配点对更加稳定。

SIFT 描述子表征图像的局部特征，可适应平移、旋转、尺度缩放、亮度的变化，并具有非常好的旋转不变性，当旋转角度从 0°到 180°时，描述子可保持 80%以上的重复度。其对光照变化、3D 视角变化、仿射变换、噪声保持一定程度的稳定性，具有较强的鲁棒性。

SIFT 描述子独特性好，信息量丰富，用于在海量特征数据库中进行匹配时，一般能获得较高的正确率。

4.3　K 均值聚类方法

聚类是无监督学习的一种方法，是常用的数据分析技术。无监督机器学习针对没有标签的情况，对样本数据进行聚类分析、关联性分析等，主要包括 K 均值聚类（K-means clustering）和关联分析。

1. 经典 K 均值聚类

如图 4.11 所示，K 均值聚类是经典的聚类方法之一，将 n 个观察对象分类到 k 个聚类，每个观察对象都被分到与均值最接近的聚类之中。

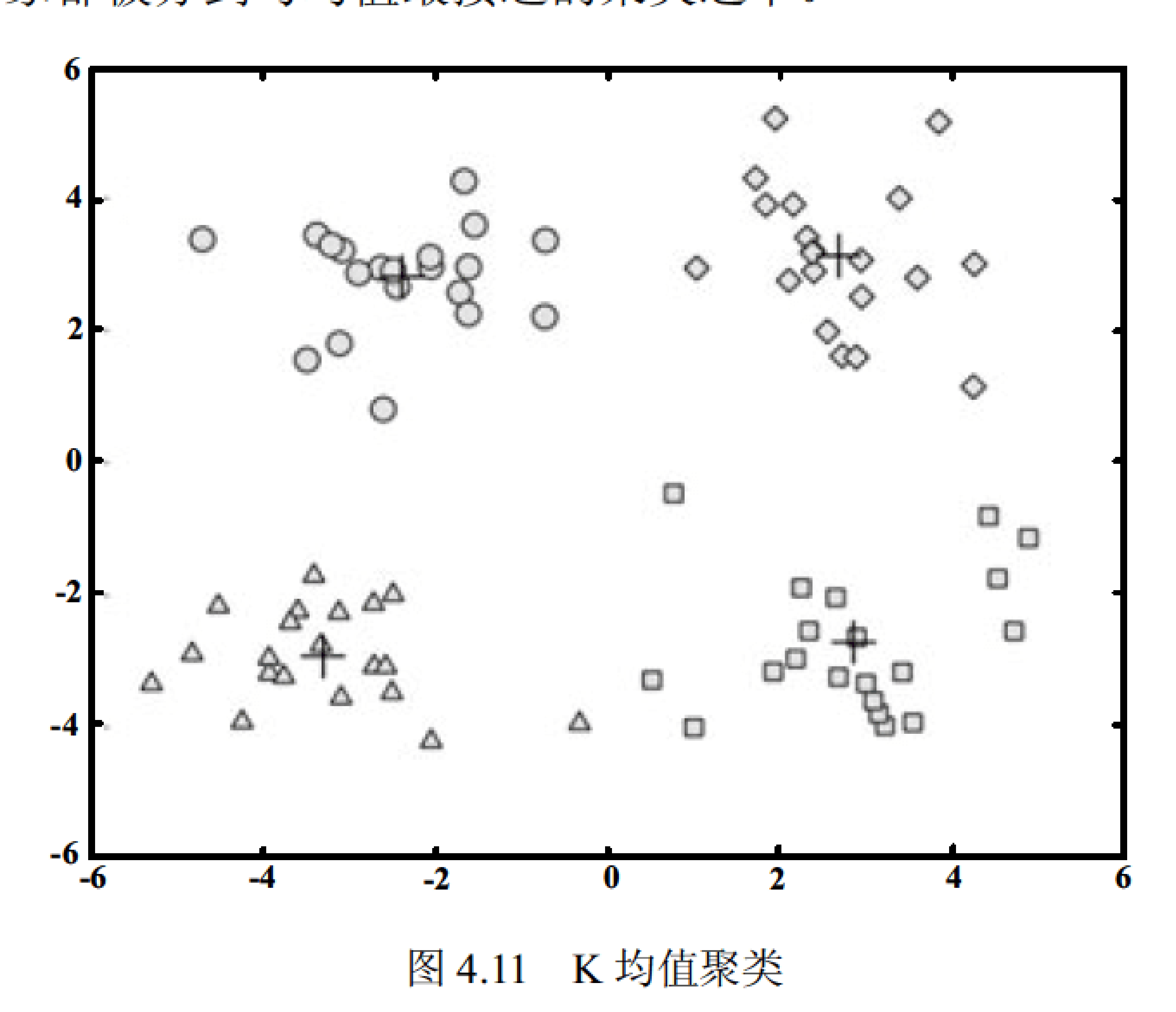


图 4.11　K 均值聚类

假如图 4.11 中是样本数据，每个样本都没有类别，明显的有 4 堆数据，用什么方法能分成 4 类呢？K 均值聚类可以解决该问题。

对象 x 隶属于集合 A 的程度可用隶属度函数表示，记做 μA(x)，其自变量包括所有可能属于集合 A 的对象，值域为 0≤μA(x)≤1。μA(x)=1 表示 x 完全隶属于集合 A。针对空间 X={x}上的隶属度函数定义一个模糊集合 A，称作在论域 X={x}上的模糊子集 $\underset{\sim}{A}$ 。

对于有限数量的对象 x_1，x_2，……，x_n，模糊集合 $\underset{\sim}{A}$ 可以表示为：

$$\underset{\sim}{A} = \{(\mu_A(x_i), x_i) \mid x_i \in X\}$$

在聚类问题中，可将聚类生成的簇看作模糊集合，每个样本点隶属于模糊集合的隶属度就是[0，1]区间里的值。

把 n 个向量 xj (j=1,2,.⋯,n)分为 c 个组 Gi (i=1,2,⋯,c)，计算每组的聚类中心，使距离最小。

当以欧几里德距离作为组 j 中向量 x_k 与相应聚类中心 c_i 间的非相似性度量时，价值函数可定义为：

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\Big(\sum_{k, x_k \in G_i} \| x_k - c_i \|^2\Big)$$

J_i 的值依赖于 G_i 的几何特性和 c_i 的位置。

假如通用距离函数 $d(x_k,c_i)$代替组 i 中的向量 x_k，则相应的总价值函数可表示为：

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\Big(\sum_{k, x_k \in G_i} \mathrm{d}(x_k - c_i)\Big)$$

通常选用欧几里德距离作为向量的非相似性指标。

经过划分后的组通常用大小为 c×n 的二维隶属矩阵 $\boldsymbol{U}$ 来定义。如果第 j 个数据点 x_j 属于组 i，则 $\boldsymbol{U}$ 中的元素 u_{ij} 为 1；否则该元素取 0。一旦确定聚类中心 c_i，可导出：

$$u_{ij} = \begin{cases} 1 & \text{对每个} k \neq i, \text{如果} \left\| x_j - c_i \right\|^2 \leq \left\| x_j - c_k \right\|^2, \\ 0 & \text{其他} \end{cases}$$

如果 c_i 是 x_j 最近的聚类中心，那么 x_j 属于组 i。由于给定的一个数据只能属于一个组，所以隶属矩阵 U 具有如下性质：

$$\sum_{i=1}^{c} u_{ij} = 1 \qquad \forall j = 1, \cdots, n$$

且

$$\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}=n$$

如果固定 u_{ij}，则使 J 最小的最佳聚类中心就是组 i 中所有向量的均值：

$$c_i=\frac{1}{|G_i|}\sum_{k,x_k\in G_i}x_k$$

$|G_i|$是 G_i 的模值或：

$$|G_i|=\sum_{j=1}^{n}u_{ij}$$

K 均值就是更新质心、更新每个样本的所属类别。

假设数据集为 $x_i\,(i=1,2,\cdots,n)$，K 均值算法重复使用下列步骤，确定聚类中心 c_i 和隶属矩阵 $\boldsymbol{U}$：

步骤01 初始化聚类中心 $c_i\,(i=1,2,\cdots,c)$，从数据点中任取 c 个点。

步骤02 确定隶属矩阵 $\boldsymbol{U}$。

步骤03 计算价值函数 J，如果它小于某个确定的阈值，或变化量小于某个阈值，可视作稳定，则迭代停止。

步骤04 修正聚类中心，返回 *Step*2。

对于给定的数据点 x，最近的聚类中心 c_i 采用下式修正：

$$\Delta c_i=\eta(x-c_i)$$

该修正公式嵌入无监督学习神经元网络的学习法则。

2. 二分 K 均值聚类

针对 K 均值聚类容易陷入局部最小的问题，有学者提出二分 K 均值聚类算法，首先把所有样本作为一个簇，然后二分该簇，接着选择其中一个簇继续进行二分。选择哪一个簇二分的原则就是能否使得误差平方和（Sum of Squared Error，SSE）尽可能小。

图 4.12 是 K 均值算法在随机初始化不好的情况下聚类的效果。

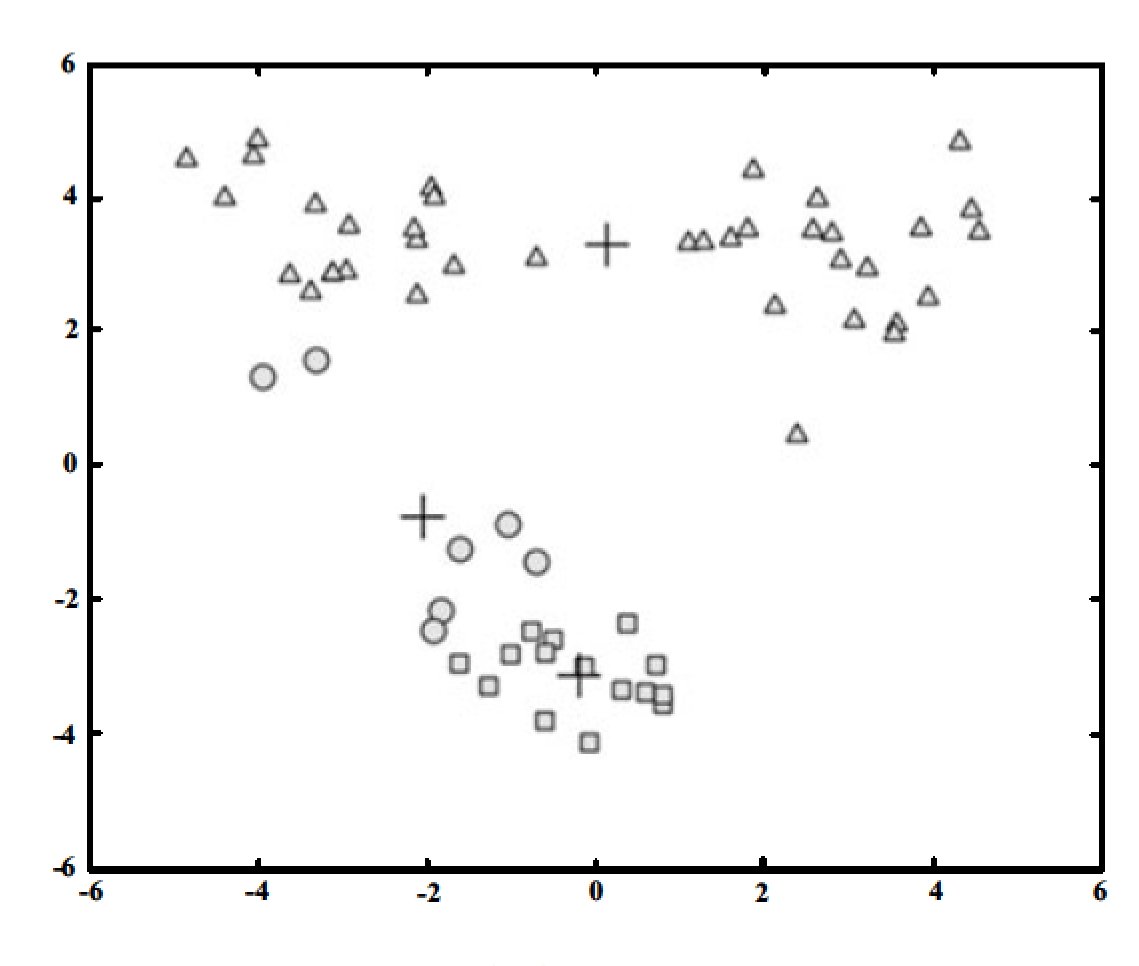


图 4.12　随机初始化不好的情况

采用二分 K 均值聚类得到的效果图如图 4.13 所示。

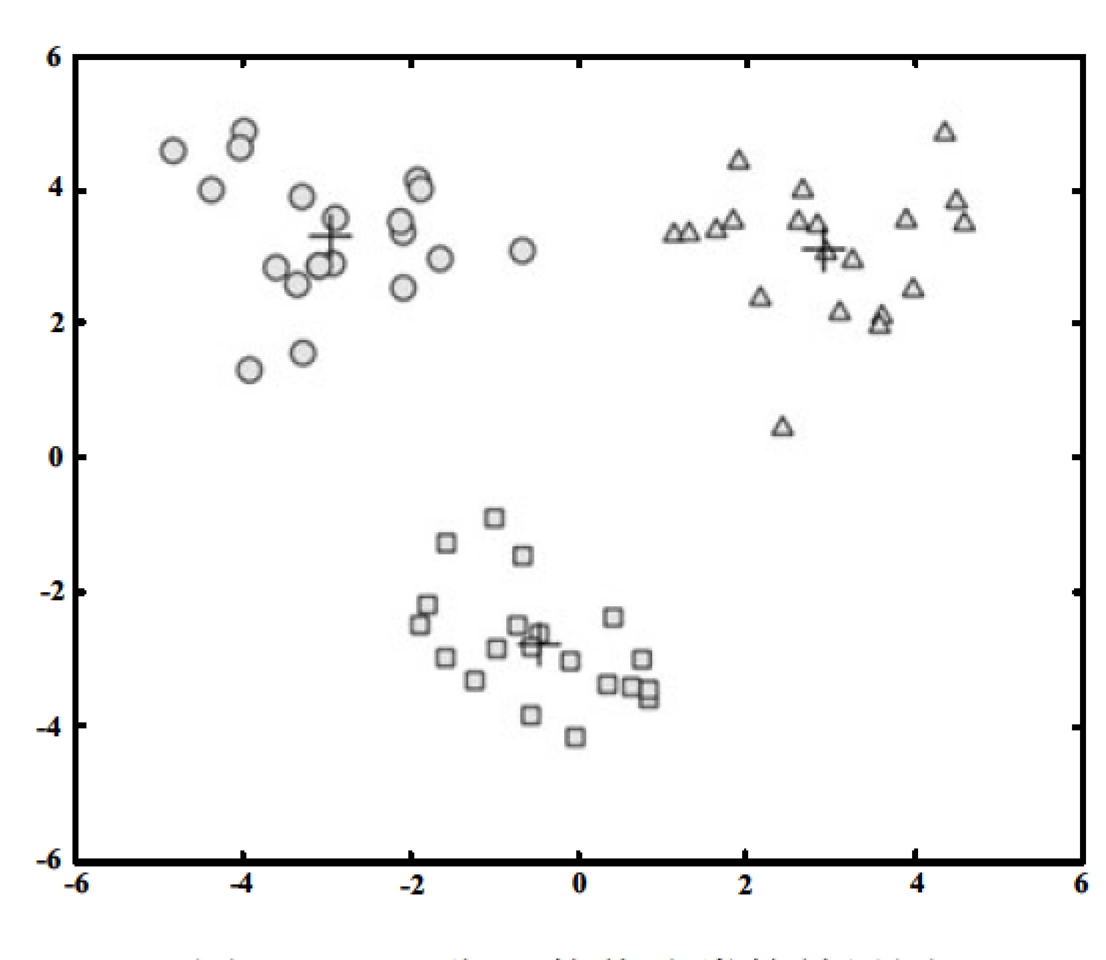


图 4.13　二分 K 均值聚类的效果图

3. 算法特点

K 均值算法简洁快速，假设均方误差是计算群组分散度的最佳参数，对于满足正态分布的数据聚类效果很好，可应用于机器学习、数据挖掘、模式识别、图像分析和生物信息学等。

K 均值算法的性能依赖于聚类中心的初始位置，不能确保收敛于最优解，对孤立点敏感。为了对其进行改善，可基于先验知识或预处理首先确定较好的初始聚类中心，或者每次随机选择不同的初始聚类中心，多次运行该算法，然后选择最优结果。

虽然二分 K 均值聚类算法改进了 K 均值聚类算法的不足，但是它们的共同的缺点

就是必须事先确定 K 的值，不合适的 K 可能返回较差的结果。对于海量的现实数据，如何确定 K 的值是学术界一直在研究的问题，常用方法是层次聚类（Hierarchical Clustering），或者借鉴 LDA 分析。

4.4 K 近邻法

机器学习分两大类，有监督学习(supervised learning)和无监督学习(unsupervised learning)。有监督学习又可分两类，即分类（classification）和回归（regression），分类的任务就是把一个样本划为某个已知类别，每个样本的类别信息在训练时需要给定，比如人脸识别、行为识别、目标检测等都属于分类。回归的任务则是预测一个数值，比如给定房屋市场的数据（面积、位置等样本信息）来预测房价走势。而无监督学习也可以成两类，即聚类（clustering）和密度估计（density estimation），聚类则是把一堆数据聚成若干组，没有类别信息；密度估计则是估计一堆数据的统计参数信息来描述数据，比如深度学习的 RBM。

K 近邻法是有监督学习方法，原理很简单，假设有一堆分好类的样本数据，分好类表示每个样本都对应一个已知类标签，当一个测试样本要我们判断它的类别时，就分别计算到每个样本的距离，然后选取离测试样本最近的前 K 个样本的标签累计投票，得票数最多的那个标签就为测试样本的标签。

图 4.14 中横坐标表示一部电影中的打斗统计个数，纵坐标表示接吻次数。对图 4.14 中的电影进行分类，其统计数据和类别如图 4.15 所示：

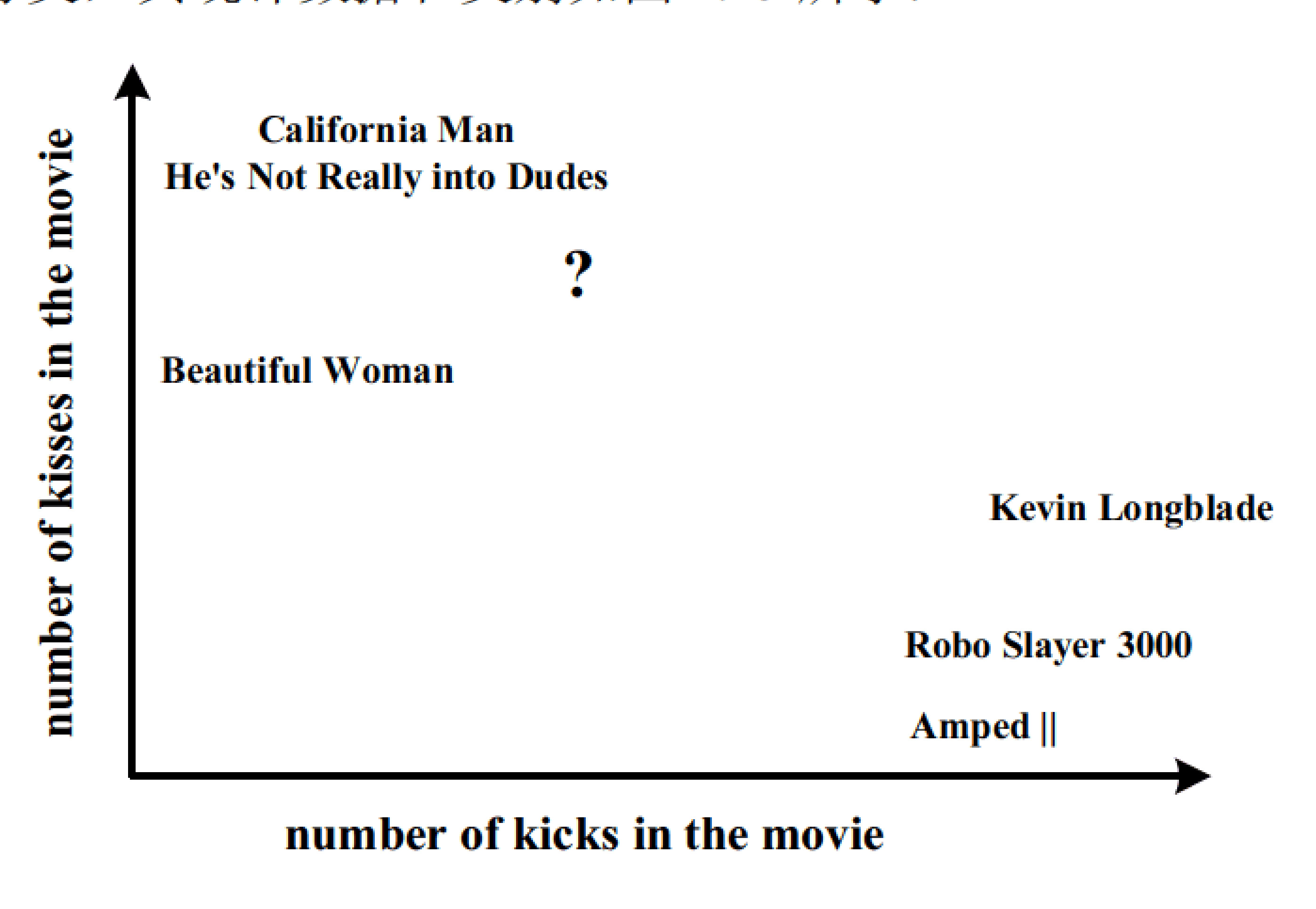


图 4.14 打斗统计

电影名称	打斗次数	接吻次数	电影类别
California Man	3	104	Romance
He's Not Really into Dudes	2	100	Romance
Beautiful Woman	1	81	Romance
Kevin Longblade	101	10	Action
Robo Slayer 3000	99	5	Action
Amped ‖	98	2	Action
?	18	90	Unknown

图 4.15　电影统计

从图 4.15 中可以看出有 3 部电影的类别是 Romance，有 3 部电影的类别是 Action，那么如何判断问号表示的这部电影的类别呢？根据 KNN 原理，需要在图 4.14 所示的坐标系中计算问号到所有其他电影之间的距离。计算出的欧式距离如图 4.16 所示。

电影名称	到该电影的距离
California Man	20.5
He's Not Really into Dudes	18.7
Beautiful Woman	19.2
Kevin Longblade	115.3
Robo Slayer 3000	117.4
Amped ‖	118.9

图 4.16　距离计算

由于我们的标签只有两类，假设我们选择 K=6/2=3，由于前 3 个距离最近的电影都是 Romance，那么问号表示的电影被判定为 Romance。

K 近邻法精度高，对离群点不敏感，对数据不需要假设模型。但是判定时计算量太大，需要大量内存。

4.5　SVM 方法

SVM（Support Vector Machine，支持向量机）的理论基础是美国电报电话公司贝尔实验室（AT&T Bell Labs., USA）的 Cortes、Corinna、Vapnik 和 Vladimir N.于 1995 年提出的统计学习理论，该理论方法对于小样本、非线性及高维模式识别问题具有较明显的优势，广泛应用于函数拟合、语音识别、文本分类、物体识别等。对应论文为 *Support-Vector Networks*（*Machine Learning*）。

在深度学习出现之前，SVM 一直占据着机器学习老大哥的位子。其理论很优美，

有很多改进版本，比如 latent-SVM、structural-SVM 等。

1. 基本原理

如图 4.17 所示，对于该数据集，3 个分类器满足分类要求，但是这个只是训练集，测试样本分布可能会比较散一些，各种可能都有。为了应对复杂情况，需要使线性分类器离两个数据集都尽可能远，因为这样会减少测试样本越过分类器的风险，提高检测精度，因此图 4.17(d)的分类器最佳。这种使得数据集到分类器之间的间距（Margin）最大化的思想就是 SVM 的核心思想，离分类器距离最近的样本称为支持向量。

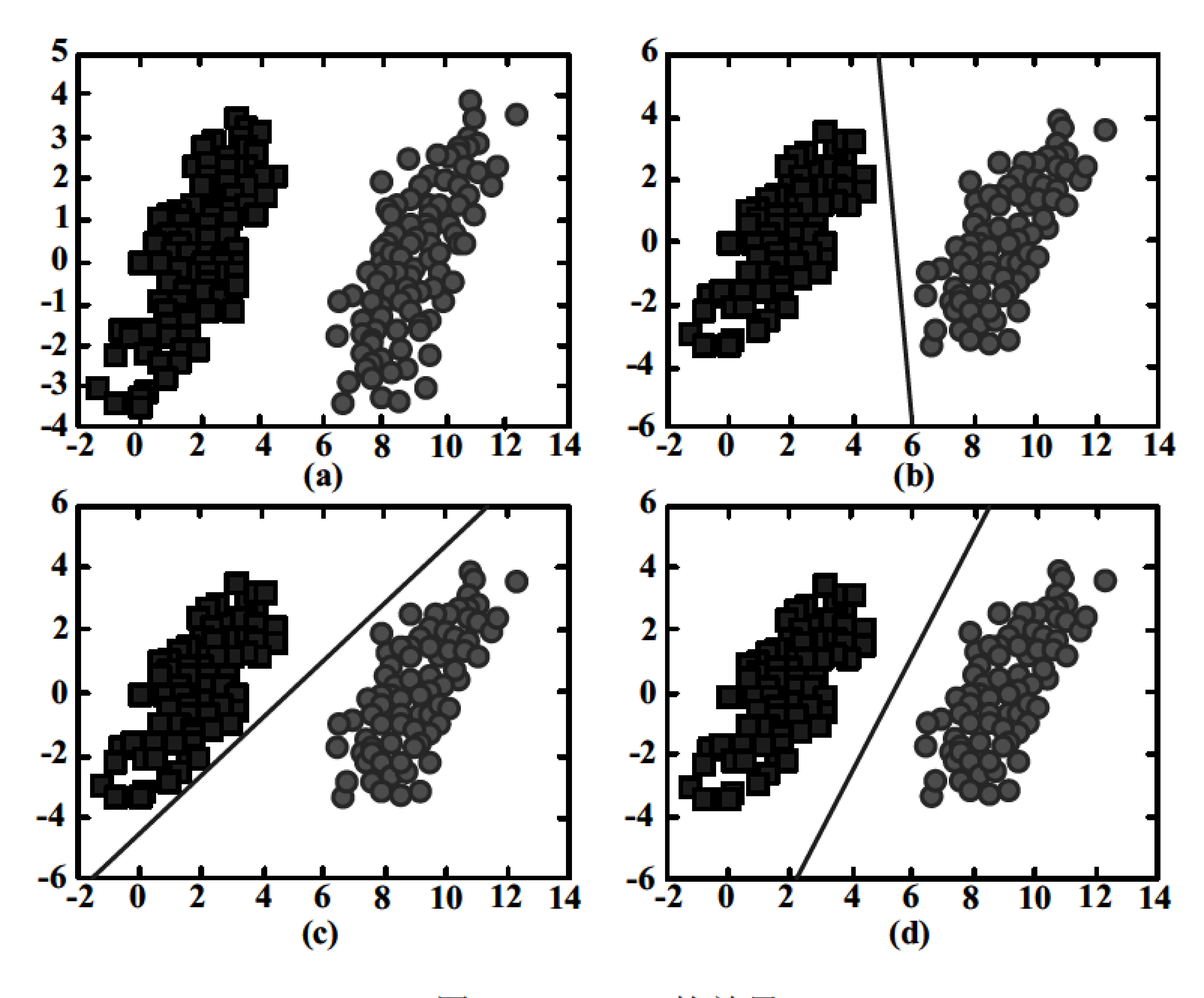


图 4.17　SVM 的效果

既然 SVM 的目标就是为了寻找最大边距，那么如何寻找支持向量？如何实现呢？

假设图 4.18 中的直线表示一个超面，为了方便观看显示成一维直线，特征都是超面维度加一维度的，特征是二维，而分类器是一维的。如果特征是三维的，分类器就是一个平面。

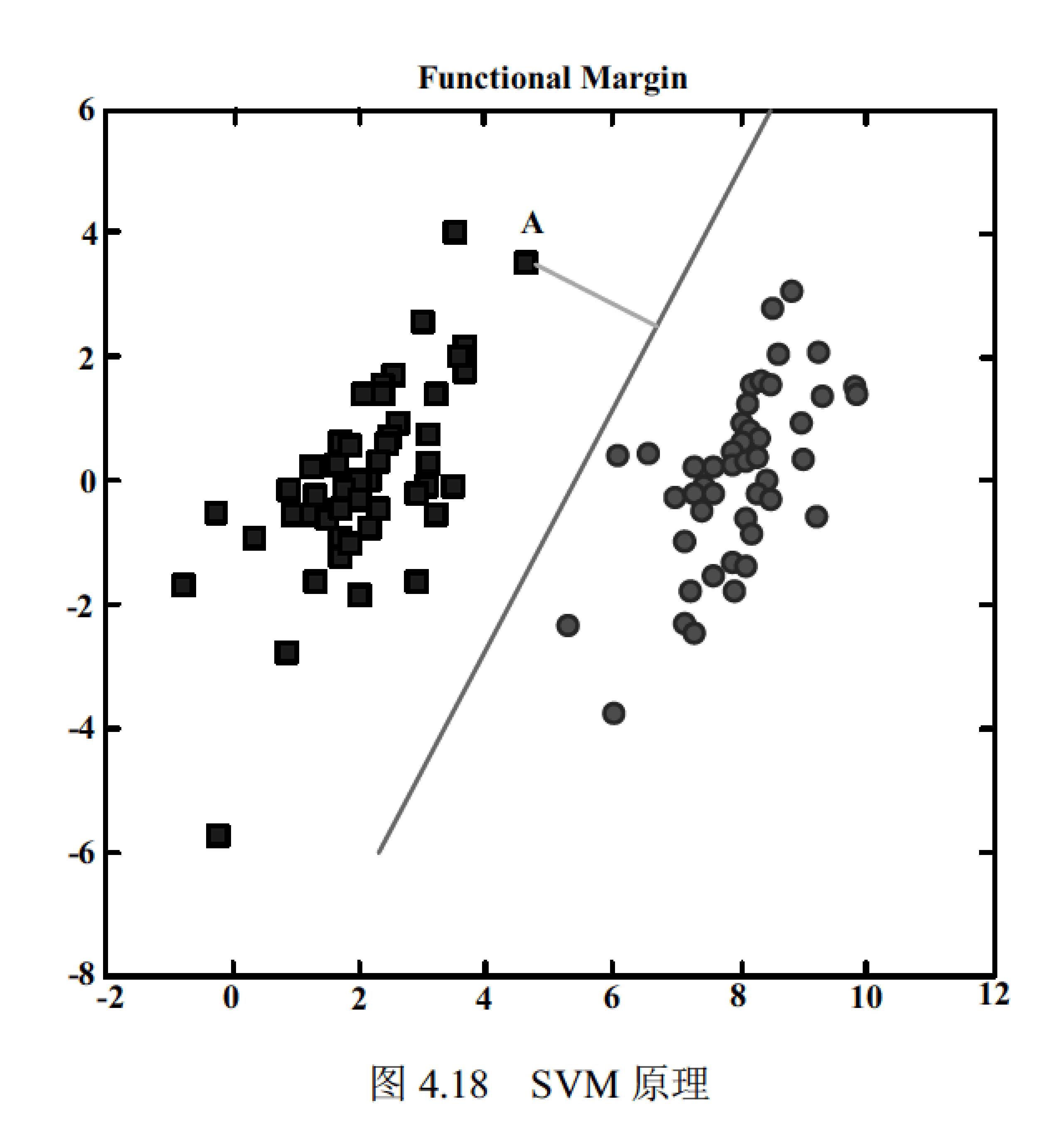


图 4.18　SVM 原理

假设超面的解析式为：

$$W^T X + b = 0$$

超面示意图如图 4.19 所示。

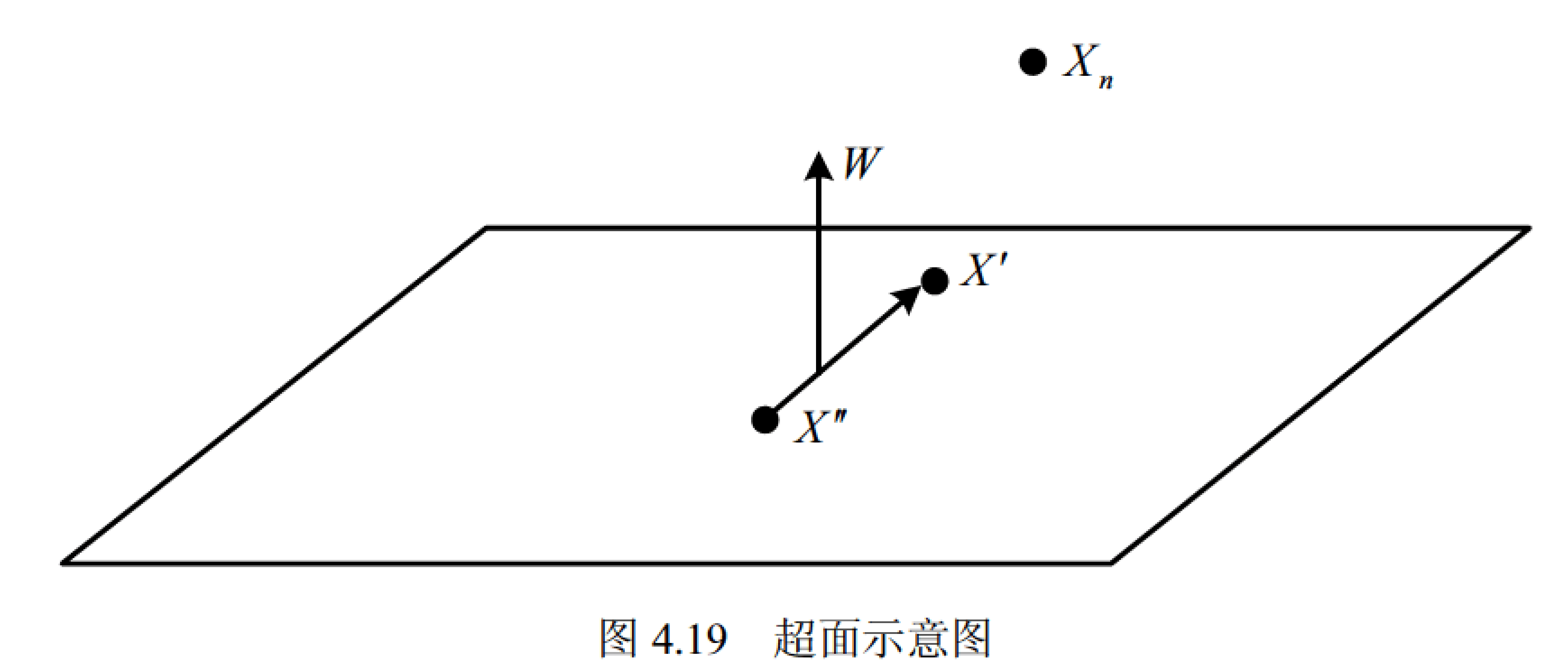


图 4.19　超面示意图

在图 4.19 中，菱形表示超面，X_n 为数据集中的一点，W 是超面权重，假设 X'和 X''是超面上的点，则：

$$\begin{cases} W^T X' + b = 0 \\ W^T X'' + b = 0 \end{cases}$$

$$\Rightarrow W^T \left(X' - X'' \right) = 0$$

因此 $\boldsymbol{W}$ 垂直于超面。那么 X_n 到超面的距离就是 X_n 和超面上任意一点 X 的连线在 $\boldsymbol{W}$ 上的投影，如图 4.20 所示。

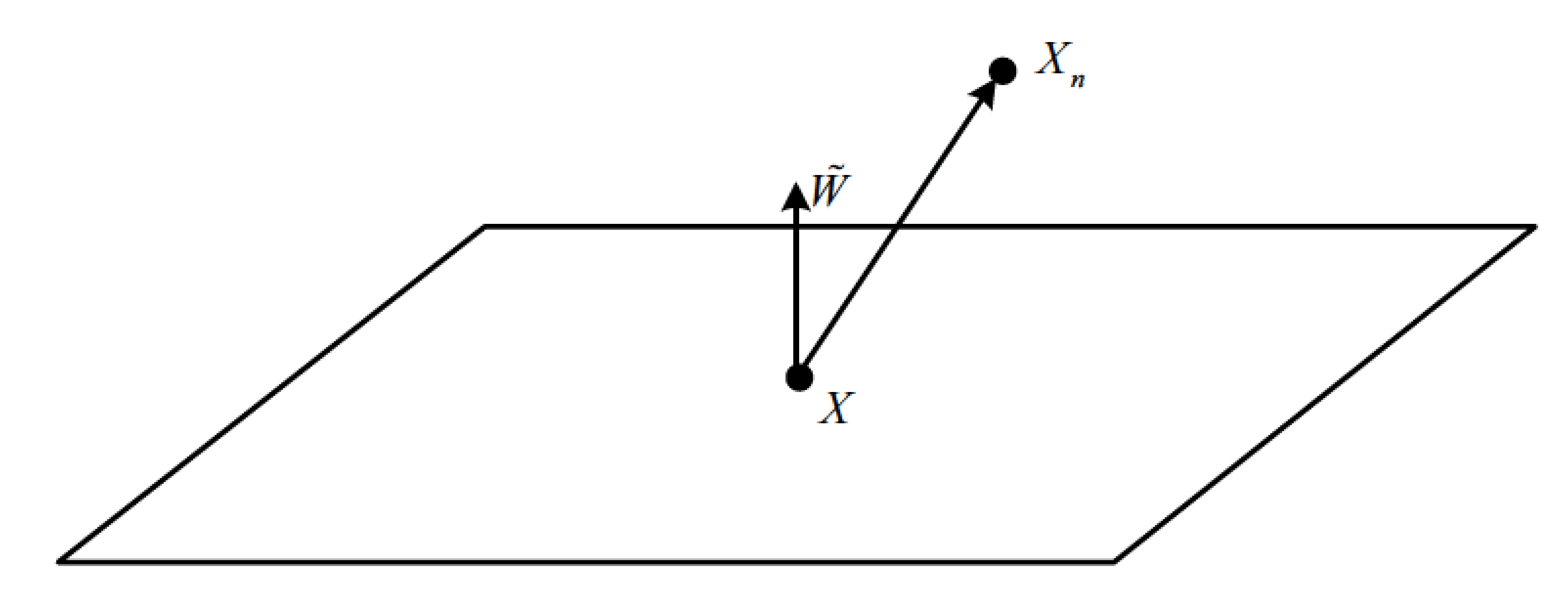


图 4.20 求取距离

$\boldsymbol{W}$ 的单位投影为：

$$\tilde{W}=\frac{W}{\|W\|}$$

（X_n-X）在 $\boldsymbol{W}$ 上的投影为：

$$\left|\tilde{W}^{\mathrm{T}}\left(X_n-X\right)\right|=\frac{1}{\|W\|}\left|W^{\mathrm{T}}X_n-W^{\mathrm{T}}X\right|=\frac{1}{\|W\|}\left|W^{\mathrm{T}}X_n+b-(W^{\mathrm{T}}X+b)\right|$$

由于点 X 位于超面之内，所以：

$$W^{\mathrm{T}}X+b=0$$

因此点 A 到超面的距离为：

$$\frac{\left|W^{\mathrm{T}}X+b\right|}{\|W\|}$$

这样可以使得分类器距所有样本距离最远，即最大化边距。但是最大化边距的前提是我们要找到支持向量，也就是离分类器最近的样本点，此时就要完成两个优化任务，找到离分类器最近的点（支持向量），然后最大化边距，即要求：

$$\underset{w,b}{\arg\max}\left\{\min_{n}\frac{\left|W^{\mathrm{T}}X_n+b\right|}{\|W\|}\right\}$$

大括号里面表示找到距离分类超面最近的支持向量，大括号外面则是使得超面离支持向量的距离最远。要优化这个函数相当困难，没有有效的优化方法。但是可以把问题

转换一下，如果把大括号里面的优化问题固定住，然后来优化外面就很容易了，可以用现在的优化方法来求解，因此我们做一个假设，假设：

$$\left|W^{\mathrm{T}}X_n+b\right|=1$$

那么只剩下优化 ***W***，整个优化公式可以写成：

$$\text{Maximize}\frac{1}{\|W\|}$$

上述过程是有等式约束的优化，约束条件为：

$$\left|W^{\mathrm{T}}X_n+b\right|=1$$

记 y_n 样本 X_n 的标签，令

$$\left|W^{\mathrm{T}}X_n+b\right|=y_n\left(W^{\mathrm{T}}X_n+b\right)$$

假设把样本 X_n 的标签设为 1 或者-1，当 X_n 在超面上面（或者右边）时，y_n 为 1，$W^{\mathrm{T}}X_n+b$ 的计算结果大于零，故 $y_n\left(W^{\mathrm{T}}X_n+b\right)$ 可以表示 x_n 离超面的距离；当 X_n 在超面下面（或者左边）时，y_n 为-1，$W^{\mathrm{T}}X_n+b$ 的计算结果小于零，故 $y_n\left(W^{\mathrm{T}}X_n+b\right)$ 仍可以表示 X_n 离超面的距离。因此把通常两类的标签 0 和 1 转换成-1 和 1，就可以把标签信息融入等式约束之中。

通常需要求解的最优化问题有如下几类。

(i) 无约束优化问题

$$\min f(x);$$

(ii) 有等式约束的优化问题

$$\min f(x),$$
$$s.t.\quad h_i(x)=0;\qquad i=1,2,\cdots,n$$

(iii) 有不等式约束的优化问题

$$\min f(x),$$
$$\begin{aligned} s.t.\quad & g_i(x)\le 0; \qquad i=1,2,\cdots,n \\ & h_j(x)=0; \qquad j=1,2,\cdots,m \end{aligned}$$

对于第(i)类的优化问题，常用 Fermat 定理，求取 $f(x)$的导数，然后令其为零，可以

求得候选最优值，再在这些候选值中验证；如果是凸函数，可以保证是最优解。

对于第(ii)类的优化问题，常用拉格朗日乘子法（Lagrange Multiplier），把等式约束 $h_i(x)$用一个系数与 $f(x)$写为一个式子，称为拉格朗日函数，而系数称为拉格朗日乘子。通过拉格朗日函数对各个变量求导，令其为零，可以求得候选值集合，然后验证求得最优解。

对于第(iii)类的优化问题，常用 KKT 条件。把所有的等式、不等式约束与 $f(x)$写为一个式子，叫拉格朗日函数，系数为拉格朗日乘子，通过一些条件，可以求出最优解的必要条件，即 KKT 条件。

SVM 问题符合第二类优化方法，最大化$||\boldsymbol{W}||$的导数可以通过最小化 $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$ 实现：

$$\text{Minimize} \quad \frac{1}{2}W^T W$$

$$s.t. \quad y_n(W^{\mathrm{T}} X_n + b) \geq 1; \qquad n = 1, 2, \cdots, N$$

上述问题可以通过拉格朗日乘子法转换为极值问题进行求解。拉格朗日函数为：

$$L(W, b, \alpha) = \frac{1}{2}W^T W - \sum_{n=1}^{N} \alpha_n \left(y_n \left(W^T X_n + b \right) - 1 \right)$$

式中 $\alpha_n \geq 0$ 为拉格朗日乘子。为了得到极值点，将拉格朗日乘子法函数分别对 $\boldsymbol{W}$ 和 b 求导，令导数为 0，得到：

$$W - \sum_{n=1}^{N} \alpha_n y_n X_n$$

$$-\sum_{n=1}^{N} \alpha_n y_n = 0$$

最终得到要求解的优化函数为：

$$\text{Minimize} \quad L(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m X_n^T X_m$$

$$s.t. \quad \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$\alpha_n \geq 0$$

现在只需要做一个二次规划即可求出 α，二次规划优化求解为：

$$\min_{\alpha}\ \frac{1}{2}\alpha^T\begin{bmatrix} y_1y_1x_1^Tx_1 & y_1y_2x_1^Tx_2 & \cdots & y_1y_Nx_1^Tx_N \\ y_2y_1x_2^Tx_1 & y_2y_2x_2^Tx_2 & \cdots & y_2y_Nx_2^Tx_N \\ \cdots & \cdots & \cdots & \cdots \\ y_Ny_1x_N^Tx_1 & y_Ny_2x_N^Tx_2 & \cdots & y_Ny_Nx_N^Tx_N \end{bmatrix}\alpha+\left(-1^T\right)\alpha$$

$$s.t.\quad \underbrace{y^{\mathrm{T}}\alpha=0}_{\text{线性约束条件}}$$

$$\underbrace{0}_{\text{下边界}}\le\alpha\le\underbrace{\infty}_{\text{上边界}}$$

在求出 α 之后，就可以求出 $\boldsymbol{W}$ 了。

到此为止，SVM 的公式推导完成，可以看出数学理论很严密，很优美。二次规划求解计算量很大，在实际应用中常用 SMO（Sequential Minimal Optimization）算法。

2. 实现过程

寻找最大化间隔的目标最终转换成求解拉格朗日乘子变量 α 的求解问题，求出 α 即可求解出 SVM 的权重 W，有了权重也就有了最大间隔距离。但是其实我们有个假设：就是训练集是线性可分的，这样求出的 α 在[0,infinite]之间。但是如果数据不是线性可分的呢？此时我们就要允许部分样本可以越过分类器，这样优化的目标函数就可以不变，只要引入松弛变量 $\xi_n\ge0$ 即可，它表示错分类样本点的代价，分类正确时它等于 0，当分类错误时，有：

$$\xi_n=\left|t_n-y\left(X_n\right)\right|$$

其中 t_n 表示样本的真实标签-1 或者 1，我们把支持向量到分类器的距离固定为 1，因此两类的支持向量间的距离肯定大于 1，当分类错误时 ξ_n 肯定大于 1，如图 4.21 所示。

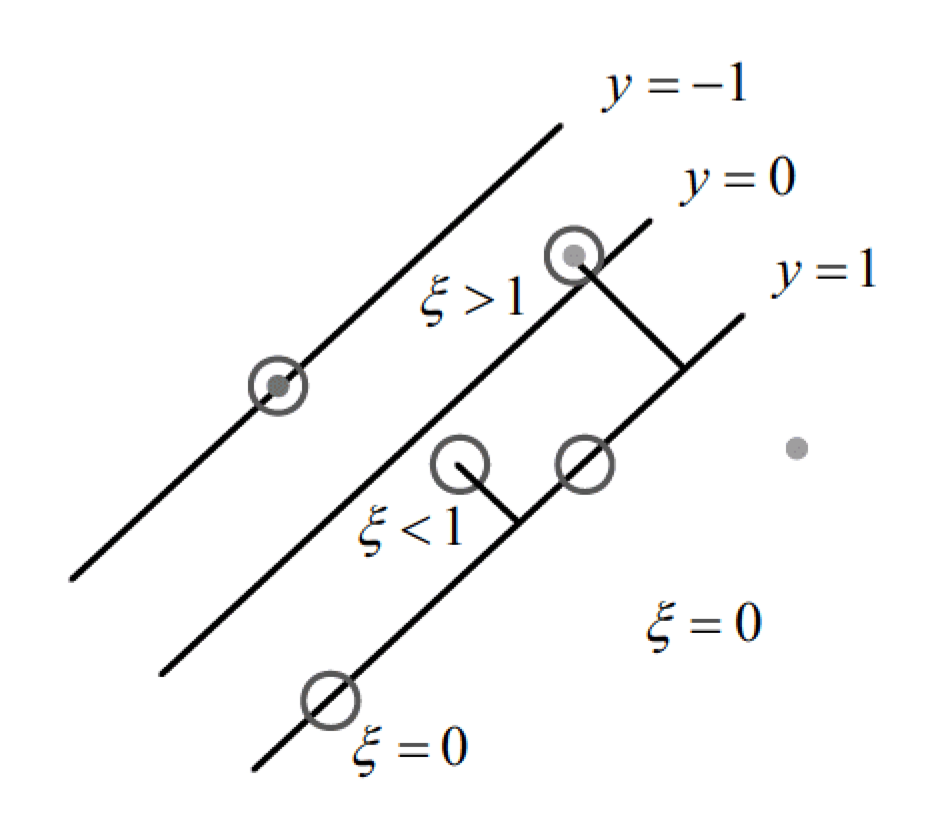


图 4.21　错分类的代价

有了错分类的代价，把目标函数添加上这一项错分类代价，得：

$$C\sum_{n=1}^{N}\xi_n+\frac{1}{2}\|W\|^2$$

采用拉格朗日乘子法，得：

$$L(W,b,a)=\frac{1}{2}\|W\|^2+C\sum_{n=1}^{N}\xi_n-\sum_{n=1}^{N}a_n\{t_n y(X_n)-1+\xi_n\}-\sum_{n=1}^{N}\mu_n\xi_n$$

多了一个 μ_n 乘子，继续求解此目标函数，求导得到：

$$\frac{\partial L}{\partial W}=0\Rightarrow W=\sum_{n=1}^{N}a_n t_n\phi(X_n)$$

$$\frac{\partial L}{\partial b}=0\Rightarrow\sum_{n=1}^{N}a_n t_n=0$$

$$\frac{\partial L}{\partial \xi_n}=0\Rightarrow a_n=C-\mu_n$$

因为 α 大于 0，μ_n 大于 0，所以 $0<\alpha<C$。

KKT 条件为：

$$a_n\geq 0$$

$$t_n y(X_n)-1+\xi_n\geq 0$$

$$a_n(t_n y(X_n)-1+\xi_n)=0$$

$$\mu_n\geq 0$$

$$\xi_n\geq 0$$

$$\mu_n\xi_n=0$$

优化函数的形式基本没变，只是多了一项错分类的价值，但是多了一个条件，即 $0<\alpha<C$，C 是一个常数，在允许有错误分类的情况下，控制最大化间距，太大会导致过拟合，太小会导致欠拟合。

这里多了一个 C 常量的限制条件，继续用 SMO 算法优化求解二次规划。

如果样本线性不可分，引入核函数后，把样本映射到高维空间就可以线性可分，如图 4.22 所示为线性不可分的样本。

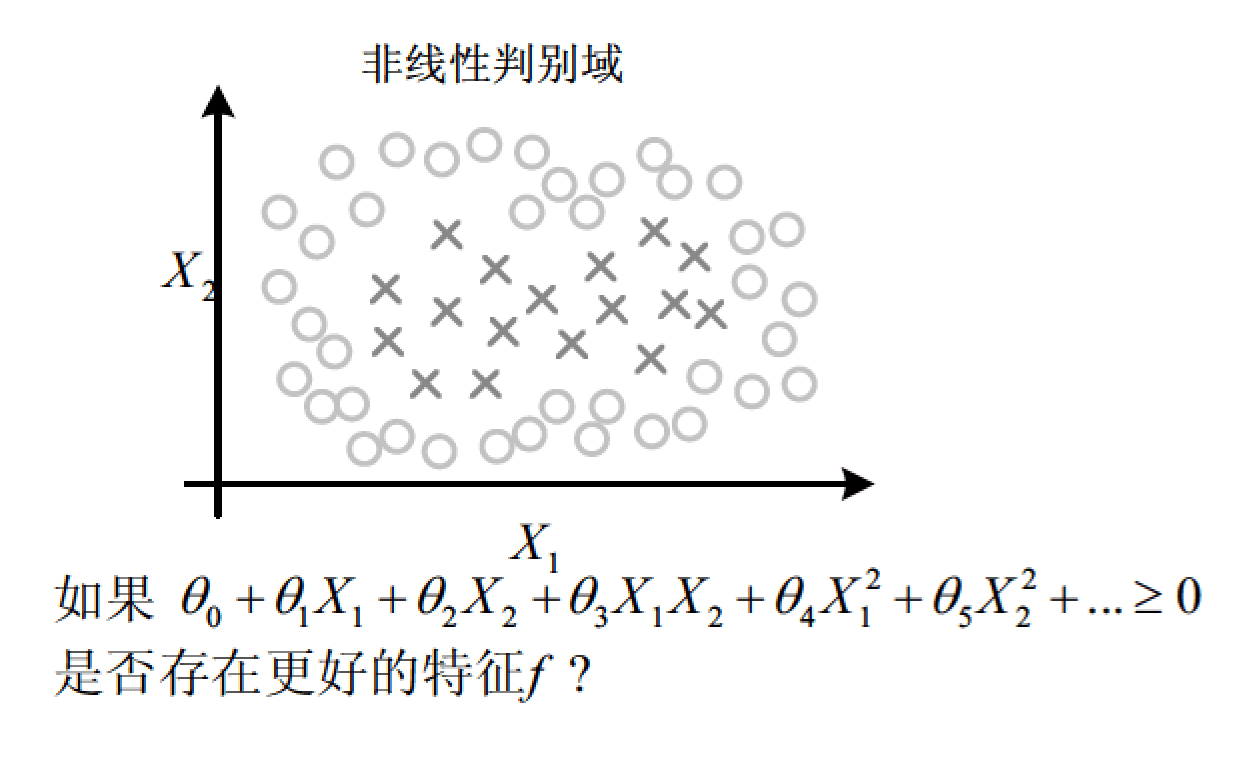


图 4.22　线性不可分的样本

在图 4.22 中，现有的样本很明显线性不可分，但是假如我们利用现有的样本 X 之间做些不同的运算，如图 4.22 右边所示的样子，让 f 作为新的样本（新的特征）是不是更好些呢？现在把 X 已经投射到高维度上去了，但是不知道 f，此时核函数就该上场了，以高斯核函数为例，选几个样本点作为基准点，利用核函数计算 f。

这样就有了 f，而核函数此时相当于对样本的 X 和基准点一个度量，做权重衰减，形成依赖于 X 的新的特征 f，把 f 放在上面说的 SVM 中继续求解 α，然后得出权重即可。

把核函数加入目标函数中：

$$\tilde{L}(a)=\sum_{n=1}^{N}a_n-\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}a_na_mt_nt_mk(x_n,x_m)$$

其中 $k(x_n,x_m)$ 是核函数，采用 SMO 优化求解即可。

3. 训练与判决

❑ SVM 训练过程

选择核函数，将训练样本映射到高维特征空间。在样本特征空间中找出各类别特征样本的最优分类超平面，得到代表各样本特征的支持向量集及其相应的 VC 可信度，形成判断各特征类别的判别函数。

❑ SVM 判决过程

如图 4.23 所示，将图像中待分类像元通过核函数映射到特征空间，作为判别函数的输入，利用分类判决函数得出分类结果。

核函数将图像各像元转换输入到判别函数之中，进行分类。

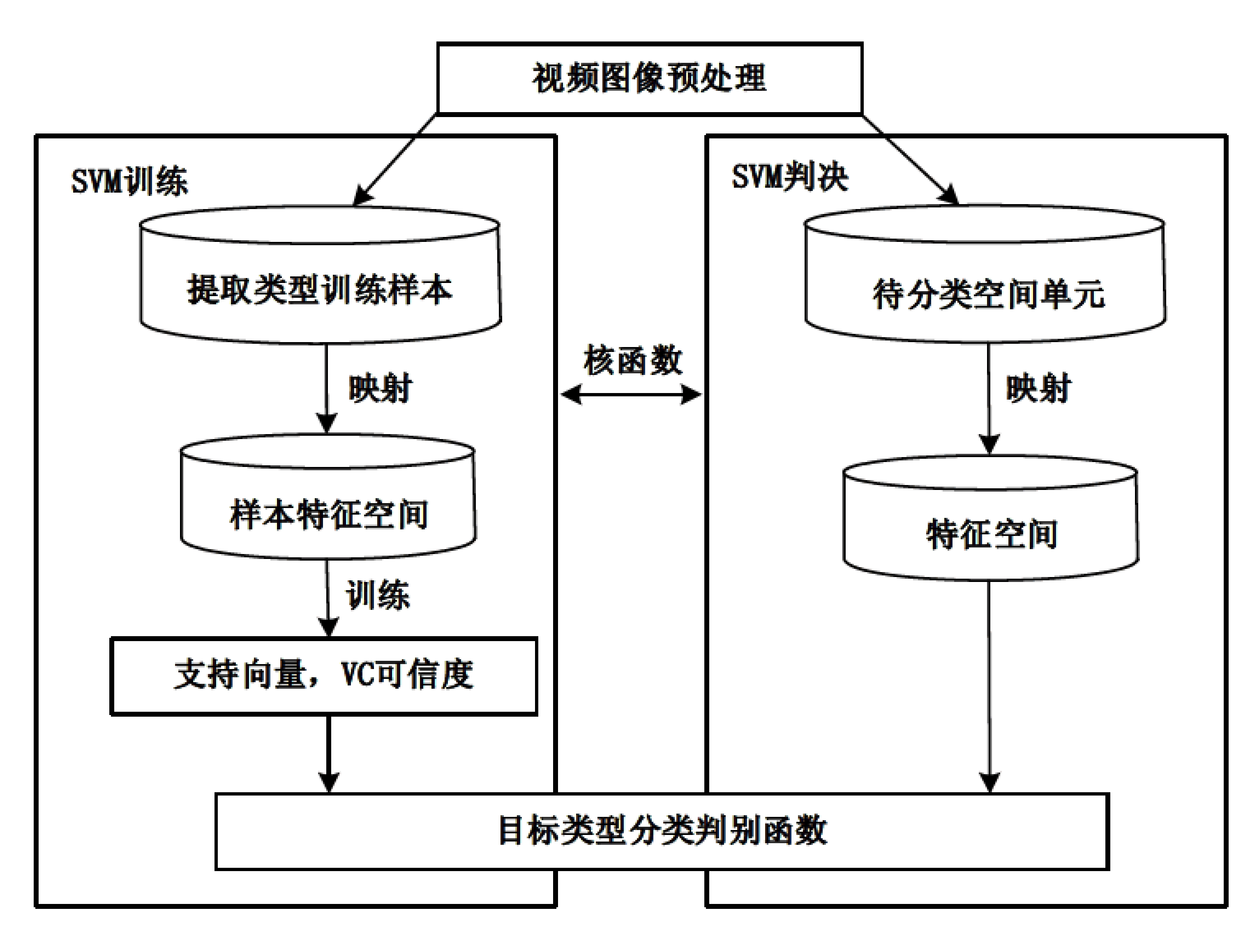


图 4.23 SVM 流程

4. 算法特点

SVM 方法建立在统计学习理论和结构风险最小的原理上，根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷。

SVM 方法的关键在于核函数，低维空间向量集通常难于划分，解决方法是将它们映射到高维空间。该办法的困难就是计算复杂度增加，核函数解决了此问题。

SVM 的最终判别函数只由少数支持向量所确定，计算的复杂性取决于支持向量的数目，而不是样本空间的维数。少数支持向量决定最终结果，不但抓住关键样本、剔除冗余样本，而且算法简单，鲁棒性好。

要建立任何一个数据模型，人为干预越少越客观。与其他方法相比，建立 SVM 模型所需要的先验干预较少。

SVM 方法对大规模训练样本难以实施，矩阵存储和计算将耗费大量的内存和运算。改进方法有 J.Platt 的 SMO 算法、T.Joachims 的 SVM、C.J.C.Burges 的 PCGC、张学工的 CSVM、O.L.Mangasarian 的 SOR。

SVM 核函数的选取以及参数确定不具有普遍性，不同的问题和区域都可能不一样，没有形成统一模式，即使最优 SVM 算法的参数选择也可能要凭借经验、实验对比获取。

SVM 方法解决多分类问题存在困难，可以通过多个二类 SVM 的组合来解决，主要有一对多组合模式、一对一组合模式和 SVM 决策树。

4.6　BP 网络

1986 年以 Rumelhart 和 McCelland 为首的科研小组提出了 BP（Back Propagation）神经网络，代表论文为 *Learning representations by back-propagating errors*。该文的主要作者 Geoffrey E. Hinton 就是深度学习提出者。

1. 基本原理

BP 神经网络是应用最广泛的神经网络模型之一，其训练方法为按误差逆传播算法，为多层前馈网络。BP 网络的特点是能学习和存储大量的输入-输出模式映射关系，而无须事前揭示描述这种映射关系的数学方程。BP 网络的学习规则为最速下降法，通过反向传播不断调整网络的权值和阈值，以网络的误差平方和最小为训练原则。

神经网络学习模型包括输入层、隐含层和输出层，典型的 BP 神经网络模型如图 4.24 所示。

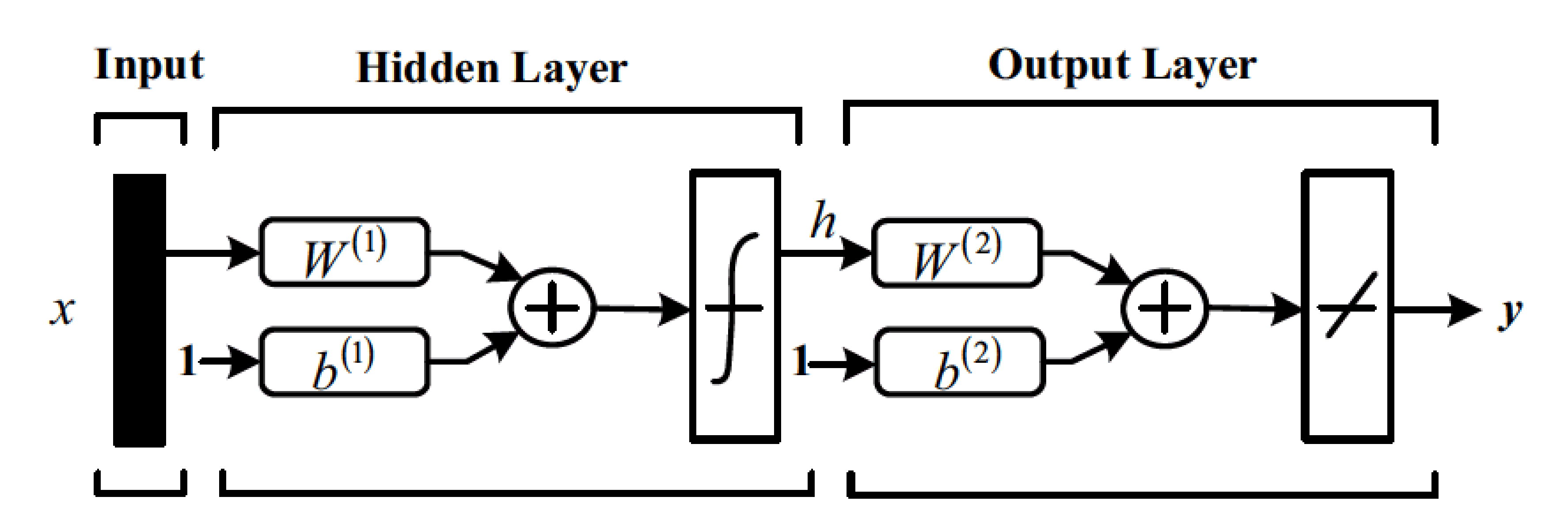


图 4.24　典型的 BP 神经网络模型

BP 神经元的传输函数为非线性函数，而在感知机中为阶跃函数，在线性神经网络中为线性函数，一般选用 log-sigmoid 函数或 tan-sigmoid 函数。BP 神经网络通常为多层神经网络，图 4.24 中所示的 BP 神经网络的隐含层的传输函数即为非线性函数，隐含层可以有多层，而输出层的传输函数不限其为线性函数或非线性函数。

输入层与隐含层的关系为：

$$h = f_1\left(W^{(1)}x + b^{(1)}\right)$$

其中 x 为 m 维特征向量（列向量），$W^{(1)}$为 $n\times m$ 维权值矩阵，$b^{(1)}$为 n 维的偏置向量（列向量）。

隐含层与输出层的关系为：

$$y = f_2\left(W^{(2)}h + b^{(2)}\right)$$

神经网络的关键之一是通过有监督的学习来确定权值。

- 学习目的：学习到一个模型，能够对输入得到期望的输出。
- 学习方式：在外界输入样本的刺激下改变网络的权值和阈值。
- 学习本质：动态调整各连接权值和阈值。
- 学习核心：连接权值和阈值的调整规则。

如图 4.25 所示，3 层 BP 神经网络的传播对象是误差，传播目的是得到所有层的估计误差，由后层误差推导前层误差。根据输出值的误差来逆向估计该层直接上一级前导层的误差，再用这个误差进一步估计更前一层的误差，如此层层逆推，获得所有层各自的误差估计。

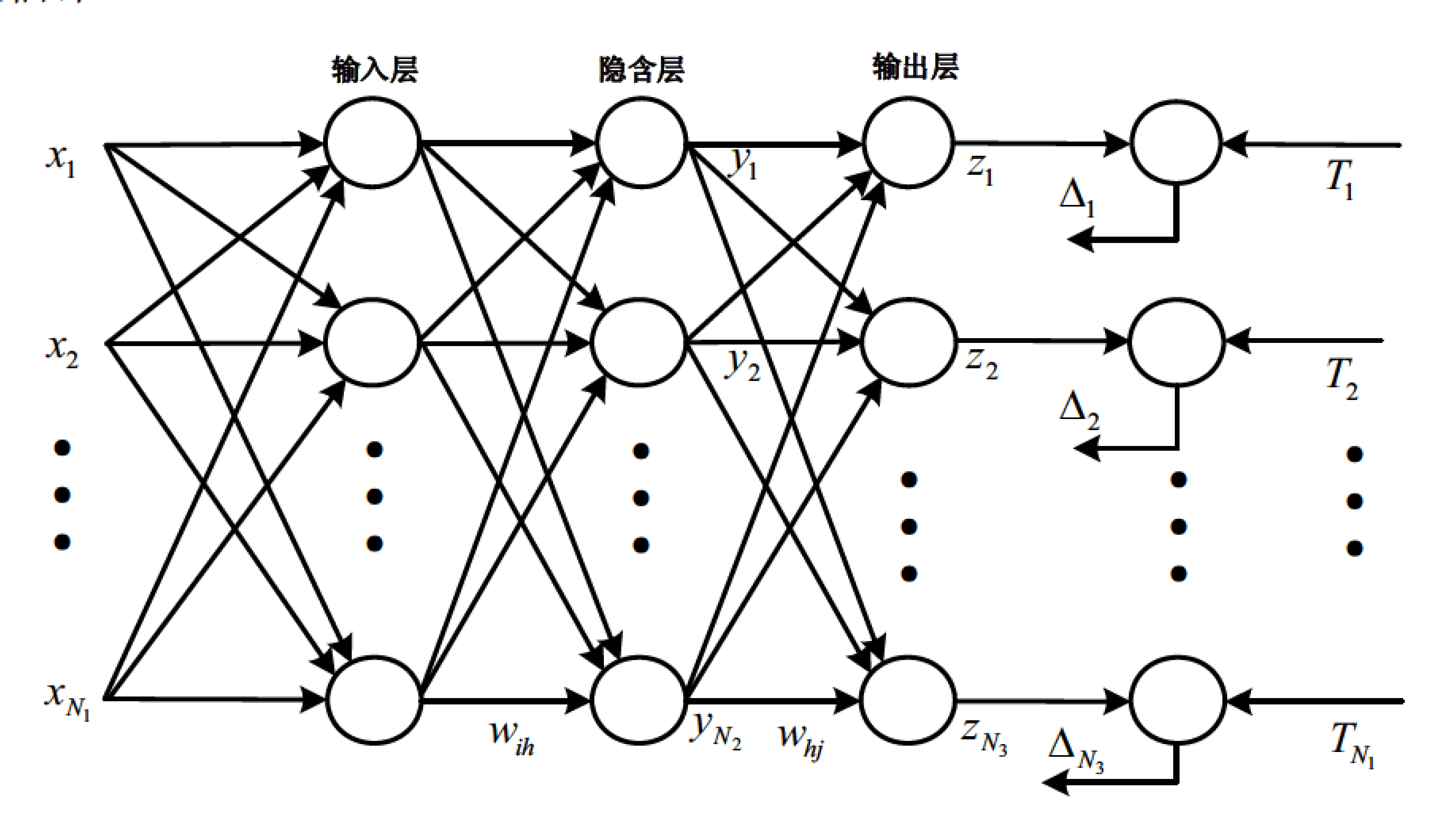


图 4.25　3 层 BP 神经网络模型

BP 利用激活函数描述层与层输出之间的关系，模拟各层神经元之间的交互反应。如图 4.26 所示，激活函数必须处处可导，常用的是 S 型函数（Simoid 或 Logistic 函数）。

输入：

$$net = x_1 w_1 + x_2 w_2 + \dots x_n w_n$$

输出：

$$y = f(net) = \frac{1}{1+e^{-net}}$$

S 型函数的导数为：

$$f'(net)=\frac{1}{1+e^{-net}}-\frac{1}{\left(1+e^{-net}\right)^2}=y(1-y)$$

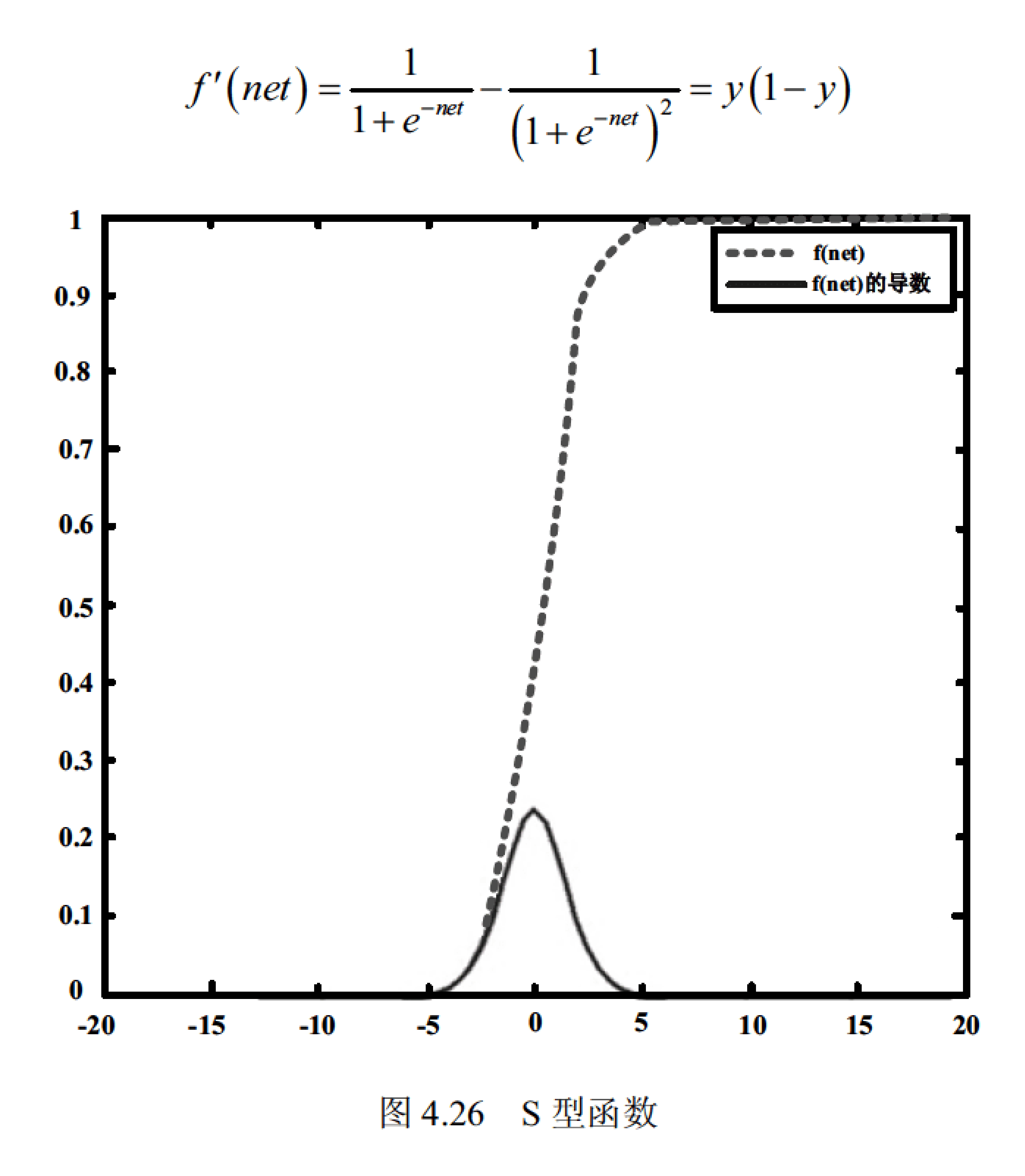


图 4.26　S 型函数

2. 实现过程

对于一个包括输入样本 X 和期望输出 Y 的训练集，BP 模型的训练过程如图 4.27 所示。其中，输入层有 n 个神经元，隐含层有 p 个神经元，输出层有 q 个神经元。输入样本为 $x=\left(x_1,x_2,\cdots,x_n\right)$；隐含层输入向量为 $hi=\left(hi_1,hi_2,\cdots,hi_p\right)$；隐含层输出向量为 $ho=\left(ho_1,ho_2,\cdots,ho_p\right)$；输出层输入向量为 $yi=\left(yi_1,yi_2,\cdots,yi_q\right)$；输出层输出向量为 $yo=\left(yo_1,yo_2,\cdots,yo_q\right)$；期望输出向量为 $d_o=\left(d_1,d_2,\cdots,d_q\right)$；输入层与中间层的连接权值为 w_{ih}；隐含层与输出层的连接权值为 w_{ho}；隐含层各神经元的阈值为 b_h；输出层各神经元的阈值为 b_o；样本数据个数为 $k=1,2\cdots,m$；激活函数为 f；误差函数为 $e=\frac{1}{2}\sum_{o=1}^{q}\left(d_o(k)-yo_o(k)\right)^2$。

正向传播：
输入样本——输入层——隐含层——输出层
判断是否转入反向传播阶段：
若输出层的实际输出与期望的输出（教师信号）不符

误差反向传播
误差以某种形式在各层表示————修正各层单元的权值

网络输出的误差减少到可接受的程度
或进行到预先设定的学习次数为止

图 4.27　BP 模型的训练过程

（1）BP 模型的实现流程

BP 模型的实现流程可分为 8 步。

步骤 01 初始化网络。

对各连接权值分别随机赋予区间(-1,1)内的初始值，设定误差函数 e，预设计算精度值 ε 和最大学习次数 M。

步骤 02 随机选取第 k 个输入样本及其期望输出。

$$x(k)=\left(x_1(k),x_2(k),\cdots,x_n(k)\right)$$

$$d_o(k)=\left(d_1(k),d_2(k),\cdots,d_q(k)\right)$$

步骤 03 计算隐含层和输出层各神经元的输入和输出。

$$hi_h(k)=\sum_{i=1}^{n}w_{ih}x_i(k)-b_h,\ h=1,2,\cdots,p$$

$$ho_h(k)=f\left(hi_h(k)\right),h=1,2,\cdots,p$$

$$yi_o(k)=\sum_{h=1}^{p}w_{ho}ho_h(k)-b_o,o=1,2,\cdots q$$

$$yo_o(k)=f\left(yi_o(k)\right),o=1,2,\cdots q$$

步骤 04 根据网络期望输出和实际输出间的差值，计算误差函数对输出层各神经元的偏导数。

$$\frac{\partial e}{\partial w_{ho}}=\frac{\partial e}{\partial yi_o}\frac{\partial yi_o}{\partial w_{ho}}$$

$$\frac{\partial yi_o(k)}{\partial w_{ho}}=\frac{\partial\left(\sum_h^p w_{ho}ho_h(k)-b_o\right)}{\partial w_{ho}}=ho_h(k)$$

$$\begin{aligned}\frac{\partial e}{\partial yi_o}&=\frac{\partial\left(\frac{1}{2}\sum_{o=1}^q\left(d_o(k)-yo_o(k)\right)\right)^2}{\partial yi_o}\\&=-\left(d_o(k)-yo_o(k)\right)yo_o'(k)\\&=-\left(d_o(k)-yo_o(k)\right)f'\left(yi_o(k)\right)\\&=\delta_o(k)\end{aligned}$$

$$\begin{aligned}\frac{\partial e}{\partial hi_h(k)}&=\frac{\partial\left(\frac{1}{2}\sum_{o=1}^q\left(d_o(k)-yo_o(k)\right)^2\right)}{\partial ho_h(k)}\frac{\partial ho_h(k)}{\partial hi_h(k)}\\&=\frac{\partial\left(\frac{1}{2}\sum_{o=1}^q\left(d_o(k)-f\left(yi_o(k)\right)\right)^2\right)}{\partial ho_h(k)}\frac{\partial ho_h(k)}{\partial hi_h(k)}\\&=\frac{\partial\left(\frac{1}{2}\sum_{o=1}^q\left(d_o(k)-f\left(\sum_{h=1}^p w_{ho}(k)-b_o\right)^2\right)\right)}{\partial ho_h(k)}\frac{\partial ho_h(k)}{\partial hi_h(k)}\\&=-\sum_{o=1}^q\left(d_o(k)-yo_o(k)\right)f'\left(yi_o(k)\right)w_{ho}\frac{\partial ho_h(k)}{\partial hi_h(k)}\\&=-\left(\sum_{o=1}^q\delta_o(k)w_{ho}\right)f'\left(hi_h(k)\right)\\&=\delta_h(k)\end{aligned}$$

步骤 05　结合输出层各神经元的 $\delta_o(k)$ 和隐含层各神经元的输出，修正连接权值 $w_{ho}(k)$。

$$\Delta w_{ho}(k)=-\mu\frac{\partial e}{\partial w_{ho}}=\mu\delta_o(k)ho_h(k)$$

$$w_{ho}^{N+1}=w_{ho}^N+\eta\delta_o(k)ho_h(k)$$

步骤06 根据隐含层各神经元的$\delta_h(k)$和输入层各神经元的输入修正连接权值。

$$\Delta w_{ih}(k) = -\mu\frac{\partial e}{\partial w_{ih}} = -\mu\frac{\partial e}{\partial hi_h(k)}\frac{\partial hi_h(k)}{\partial w_{ih}} = \delta_h(k)x_i(k)$$

$$w_{ih}^{N+1} = w_{ih}^{N} + \eta\delta_h(k)x_i(k)$$

步骤07 计算全局误差。

$$E = \frac{1}{2m}\sum_{k=1}^{m}\sum_{o=1}^{q}\left(d_o(k) - y_o(k)\right)^2$$

步骤08 当误差小于预设精度或学习次数大于设定的最大次数时，算法结束。否则，重新选取下一个学习样本及对应的期望输出，跳转到 ***Step*3**，开始下一轮的学习。

（2）BP 模型的权值调整方向

如图 4.28 所示，当误差对权值的偏导数大于零时，权值调整量为负，实际输出大于期望输出，权值向减少方向调整，使得实际输出与期望输出的差减少。

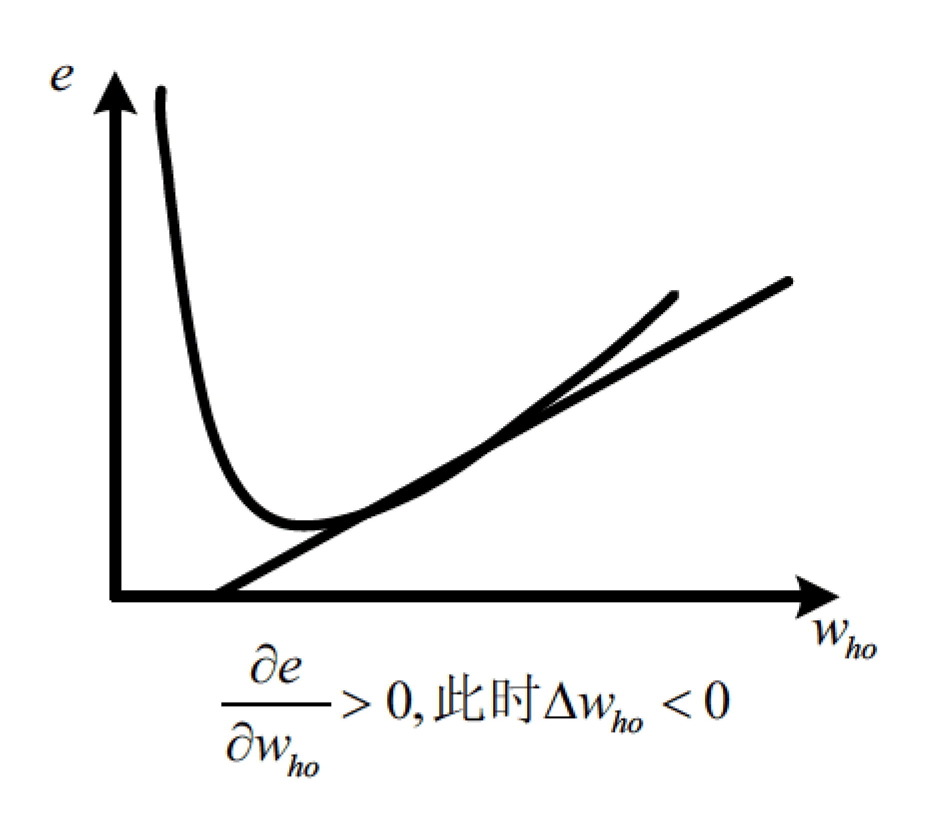


图 4.28 权值减少

如图 4.29 所示，当误差对权值的偏导数小于零时，权值调整量为正，实际输出少于期望输出，权值向增大方向调整，使得实际输出与期望输出的差减少。

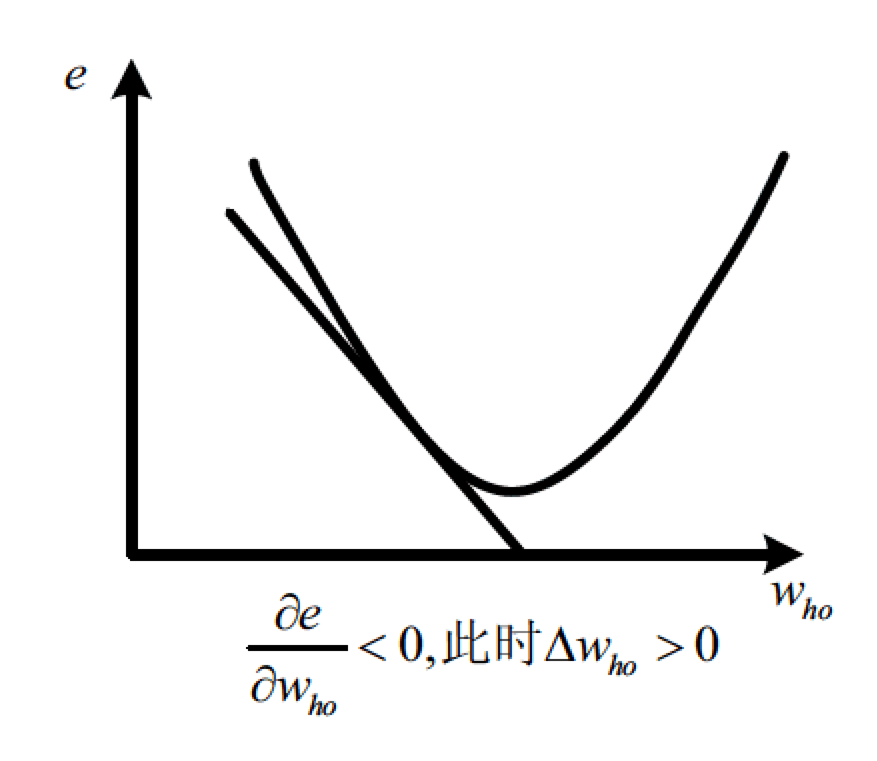


图 4.29 权值增大

BP 神经网络模型结构可分为输入层（input）、隐含层(hide layer)和输出层(output layer)。输入层神经元的个数与样本属性的维度相关，输出层神经元的数量由样本种类数决定。隐含层的层数和每层的神经元数量由用户设定。每一层包含若干个神经元，每个神经元包含一个阈值θ_j，用来调节神经元的活性。网络中的弧线 w_{ij} 表示前一层神经元和后一层神经元之间的权值。每个神经元都有输入和输出。输入层的输入和输出都是训练样本的属性值。

隐含层和输出层的输入为：

$$I_j = \sum_i w_{ij} O_i + \theta_j$$

其中，w_{ij} 是由上一层的单元 i 到单元 j 的连接权值；O_j 是上一层的单元 i 的输出；而θ_j 是单元 j 的阈值。

神经网络的学习率通常取 0 和 1 之间的值，有助于找到全局最优结果。如果学习率太小，学习过程缓慢；反之则可能会在不恰当的解之间摆动的情况。

（3）BP 算法的伪代码

BP 算法的基本流程如下。

步骤 01 初始化网络权值和神经元的阈值（如随机初始化）。

步骤 02 前向传播：逐层计算隐含层和输出层神经元的输入和输出。

步骤 03 后向传播：修正权值和阈值。

跳至步骤 2，直到满足终止条件。

BP 算法的伪代码为：

```
BPTrain()
```

```
{
   Begin：初始化 network 的权和阈值。
   while 终止条件不满足
{
      for samples 中的每个训练样本 X {
        // 向前传播输入
        for 隐藏或输出层每个单元 j {
            I_j = ∑_i w_ij O_i + θ_j； // 计算单元 j 的输入
            O_j = 1/(1+e^(-I_j))； // 计算单元 j 的输出
        }
        // 后向传播误差
        for 输出层每个单元 j {
            Err_j = O_j(1-O_j)(T_j-O_j)； // 计算误差
        }
        for 由最后一个到第一个隐含层，对于隐含层每个单元 j {
            Err_j = O_j(1-O_j)∑_k Err_k w_kj； // k 是 j 的下一层神经元
        }
        for network 中每个权值 w_ij {
            Δw_ij = (l)Err_j O_i；  // 权值增值
            w_ij = w_ij + Δw_ij；  // 权值更新
        }
        for network 中每个阈值θ_j {
            Δθ_j = (l)Err_j；  // 阈值增值
            θ_j = θ_j + Δθ_j；  // 阈值更新
        }
     }
}
```

（4）BP 模型的学习

对于输出层神经元 $Err_j = O_j\left(1-O_j\right)\left(T_j-O_j\right)$，$O_j$ 是单元 j 的实际输出，而 T_j 是 j 基于给定训练样本的已知类标号的真正输出。

对于隐含层神经元 $Err_j = O_j\left(1-O_j\right)\sum_k Err_k w_{kj}$，$w_{ij}$ 是由下一较高层中单元 k 到单元 j 的连接权，Err_k 是单元 k 的误差。

$\Delta w_{ij} = (l) Err_j O_i$ 为权值增量，$\Delta\theta_j = (l) Err_j$ 为阈值增量，其中 l 是学习率。

Err_j 是神经元的误差，对于 Err_j 的推导采用梯度下降算法，其原则是保证输出单元的均方差最小。

$$E_A=\frac{1}{2}\sum_{p=1}^{P}\sum_{l=0}^{m-1}\left(d_l^{(p)}-y_l^{(p)}\right)^2$$

其中 P 是样本总数，m 是输出层神经元个数，$d_l^{(p)}$ 是样本实际输出，$y_l^{(p)}$ 是神经网络输出。

对 E_A 求 w_{ij} 的导数，实现梯度下降

如图 4.30 所示，对于输出层：

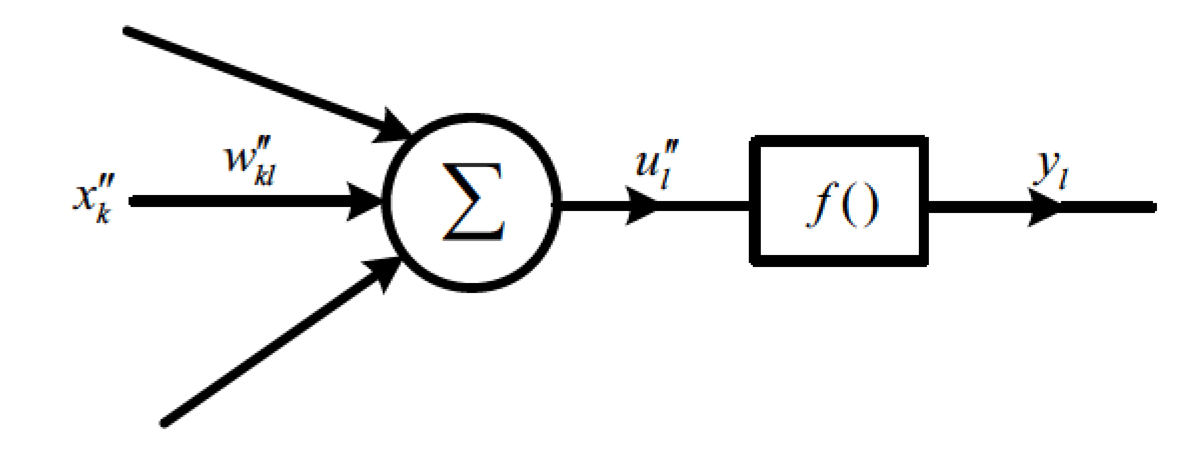


图 4.30　输出层

$$\frac{\partial E_A}{\partial w_{kl}''}=\sum_{p=1}^{P}\frac{\partial E^{(p)}}{\partial w_{kl}'}=\sum_{p=1}^{P}\frac{\partial E^{(p)}}{\partial y_l^{(p)}}\times\frac{\partial y_l^{(p)}}{\partial u_l''^{(p)}}\times\frac{\partial u_l'^{(p)}}{\partial w_{kl}''}$$

$$\because u_l''^{(p)}=\sum_k w_{kl}''\times x_k''^{(p)}$$

$$y_l^{(p)}=1/\left(1+e^{-u_l'^{p}}\right)$$

$$E^{(P)}=\frac{1}{2}\sum_l\left(d_l^{(p)}-y_l^{(p)}\right)^2$$

$$\therefore\frac{\partial E_A}{\partial w_{kl}'}=-\sum_p\left(\left(d_l^{(p)}-y_l^{(p)}\right)\times\left(y_l^{(p)}\times\left(1-y_l^{(p)}\right)\right)\times x_k''^{(p)}\right)$$

其中，$\left(d_l^{(p)}-y_l^{(p)}\right)y_l^{(p)}\left(1-y_l^{(p)}\right)$ 就是 $Err_j=O_j\left(1-O_j\right)\left(T_j-O_j\right)$。

如图 4.31 所示，对于隐含层：

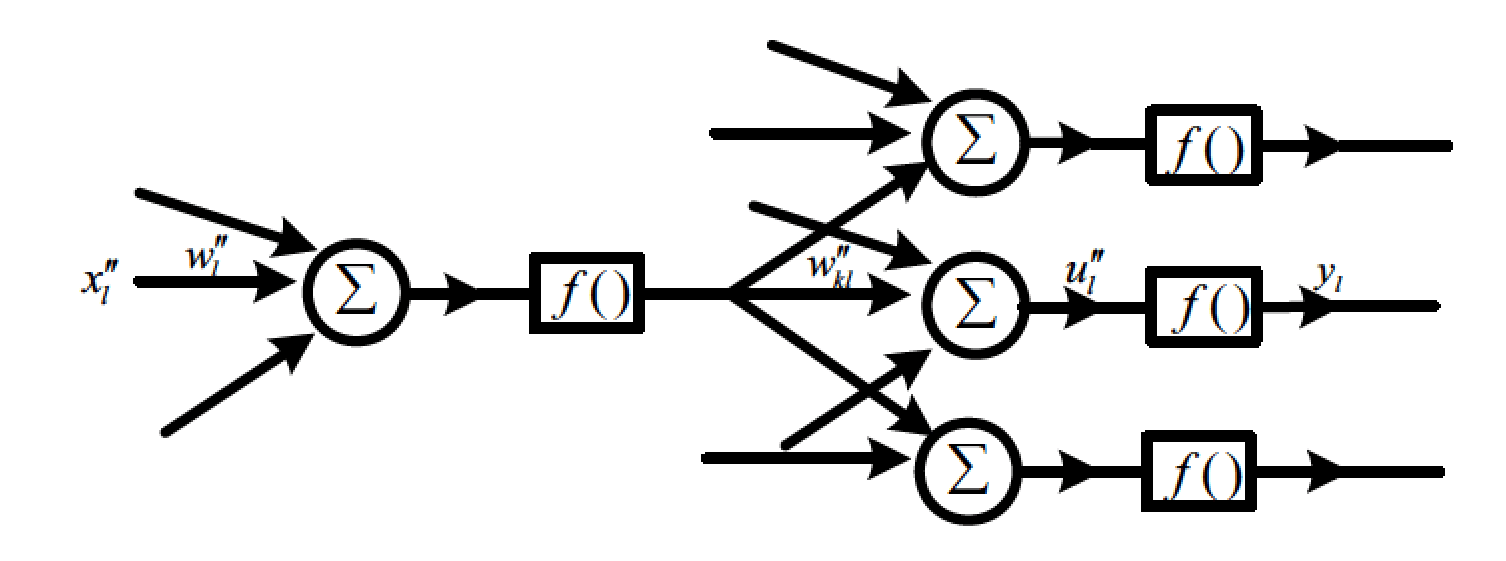


图 4.31　隐含层

$$\frac{\partial E_A}{\partial w'_{kl}}=\sum_{p=1}^{p}\frac{\partial E^{(p)}}{\partial w''_{kl}}=\sum_{p=1}^{p}\sum_{l=1}^{m-1}\frac{\partial E^{(p)}}{\partial y_l^{(p)}}\times\frac{\partial y_l^{(p)}}{\partial u_l''^{(p)}}\times\frac{\partial u_l''^{(p)}}{\partial x_k''}\times\frac{\partial x_k''}{\partial u_{kp}''}\times\frac{\partial u_{kp}''^{(p)}}{\partial w_{kl}''}$$

$$\frac{\partial E_A}{\partial w'_{kl}}=-\sum_{p=1}^{p}\sum_{l=1}^{m-1}\left(d_l^{(p)}-y_l^p\right)f'\left(u_l''^{(p)}\right)w_{kl}''x_k''^{(p)}\left(1-x_k''^{(p)}\right)x_j'^{(p)}$$

$$\frac{\partial E_A}{\partial w'_{kl}}=-\sum_{p=1}^{p}\sum_{l=1}^{m-1}\delta_{kl}^{(p)}w_{kl}''x_k'^{(p)}\left(1-x_k''^{(p)}\right)x_j'^{(p)}$$

$$\frac{\partial E_A}{\partial w'_{kl}}=-\sum_{p=1}^{p}\delta_{jk}^{(p)}x_j'^{(p)}$$

其中，$\delta_{jk}^{(p)}=\sum_{l=1}^{m-1}\delta_{kl}^{(p)}w_{kl}''x_k'^{(p)}\left(1-x_k''^{(p)}\right)$就是隐含层的误差计算公式。

（5）BP 网络设计

BP 网络设计包括网络层数、每层中的神经元个数和激活函数、初始值以及学习速率等。

理论证明，任何有理函数可由具有偏差和至少一个 S 型隐含层加上一个线性输出层的网络来逼近实现。增加层数可以进一步降低误差，提高精度，但同时也使网络复杂化。不能用仅具有非线性激活函数的单层网络来解决问题，因为能用单层网络解决的问题，用自适应线性网络也一定能解决，而且自适应线性网络的运算速度更快，而对于只能用非线性函数解决的问题，单层精度又不够高，也只有增加层数才能达到期望的结果。

网络训练精度的提高，可以通过采用一个隐含层，而增加其神经元个数的方法来获得，这在结构实现上要比增加网络层数简单得多。用精度和训练网络的时间来衡量一个神经网络设计的好坏：神经元数太少时，网络不能很好地学习，训练迭代的次数也比较多，训练精度也不高。神经元数太多时，网络的功能越强大，精确度也更高，训练迭代的次数也大，可能会出现过拟合现象。神经网络隐层神经元个数的选取原则是：在能够解决问题的前提下，再加上一两个神经元，以加快误差下降速度即可。

学习速率一般选取为 0.01～0.8，大的学习速率可能导致系统的不稳定，小的学习速率导致收敛太慢，需要较长的训练时间。对于较复杂的网络，在误差曲面的不同位置可能需要不同的学习速率，为了减少寻找学习速率的训练次数及时间，比较合适的方法是采用变化的自适应学习速率，使网络在不同的阶段设置不同大小的学习速率。

在设计网络的过程中，期望误差值也应当通过对比训练后确定一个合适的值，这个合适的值是相对于所需要的隐含层节点数来确定的。可以同时对两个不同的期望误差值的网络进行训练，最后通过综合因素来确定其中一个网络。

3. 算法特点

BP 算法需要较长的训练时间，主要由于学习速率太小而造成，可采用变化的或自适应的学习速率来加以改进。

完全不能训练，主要表现在网络的麻痹上，通常为了避免这种情况的产生，一是选取较小的初始权值，二是采用较小的学习速率。

采用的梯度下降法可能收敛到局部最小值，采用多层网络或较多的神经元，有可能得到更好的结果。

BP 算法改进的主要目标是加快训练速度、避免陷入局部极小值等，常见的改进方法有带动量因子算法、自适应学习速率、变化的学习速率以及作用函数后缩法等。动量因子法是在反向传播的基础上，在每一个权值的变化上加上一项正比于前次权值变化的值，并根据反向传播法来产生新的权值变化。自适应学习速率方法只针对一些特定的问题。改变学习速率方法的原则是，在连续几次迭代中，若目标函数对某个权倒数的符号相同，则这个权的学习速率增加，反之若符号相反，则学习速率减小。而作用函数后缩法则是将作用函数进行平移，即加上一个常数。

4.7　多感知器模型

神经网络模型如图 4.32 所示。

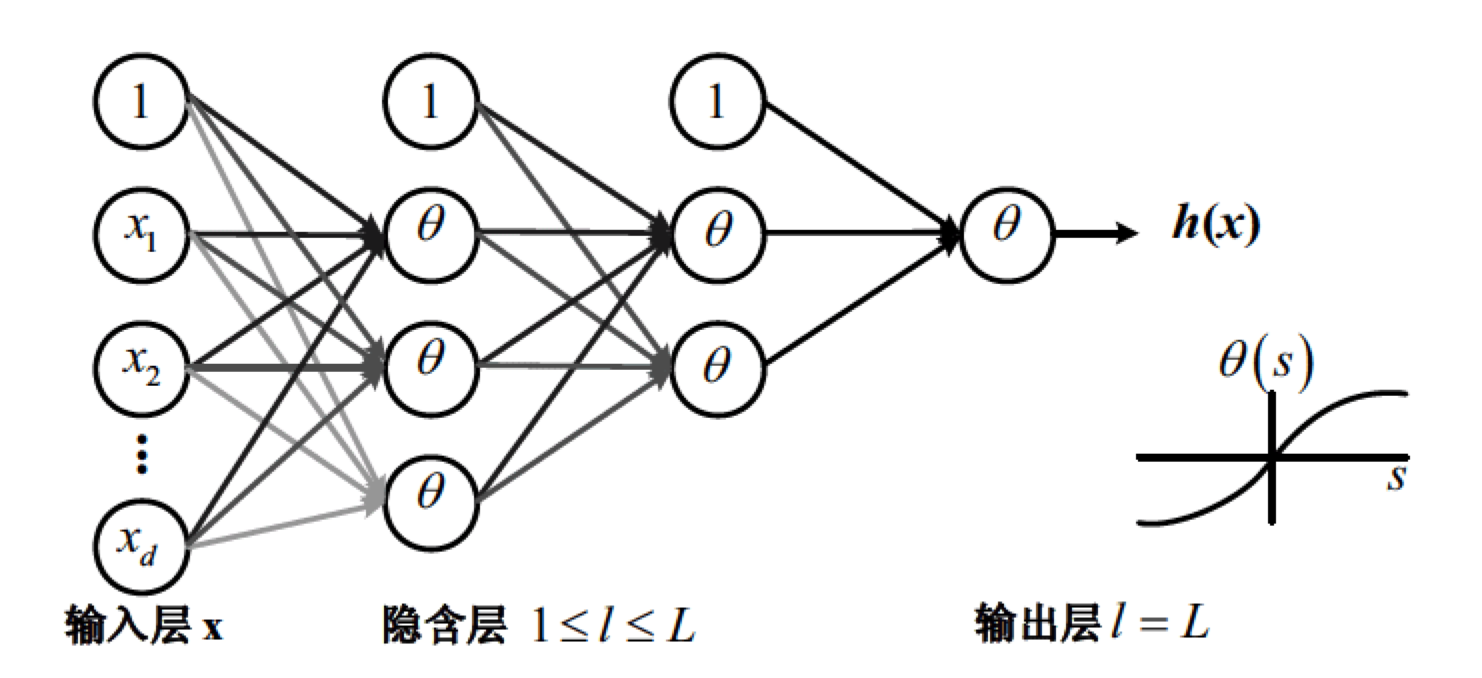


图 4.32　神经网络模型

图 4.32 中的神经网络模型由多个感知器（Perceptron）分几层组合而成，感知器就是单层的神经网络，只有一个输出节点，如图 4.33 所示。

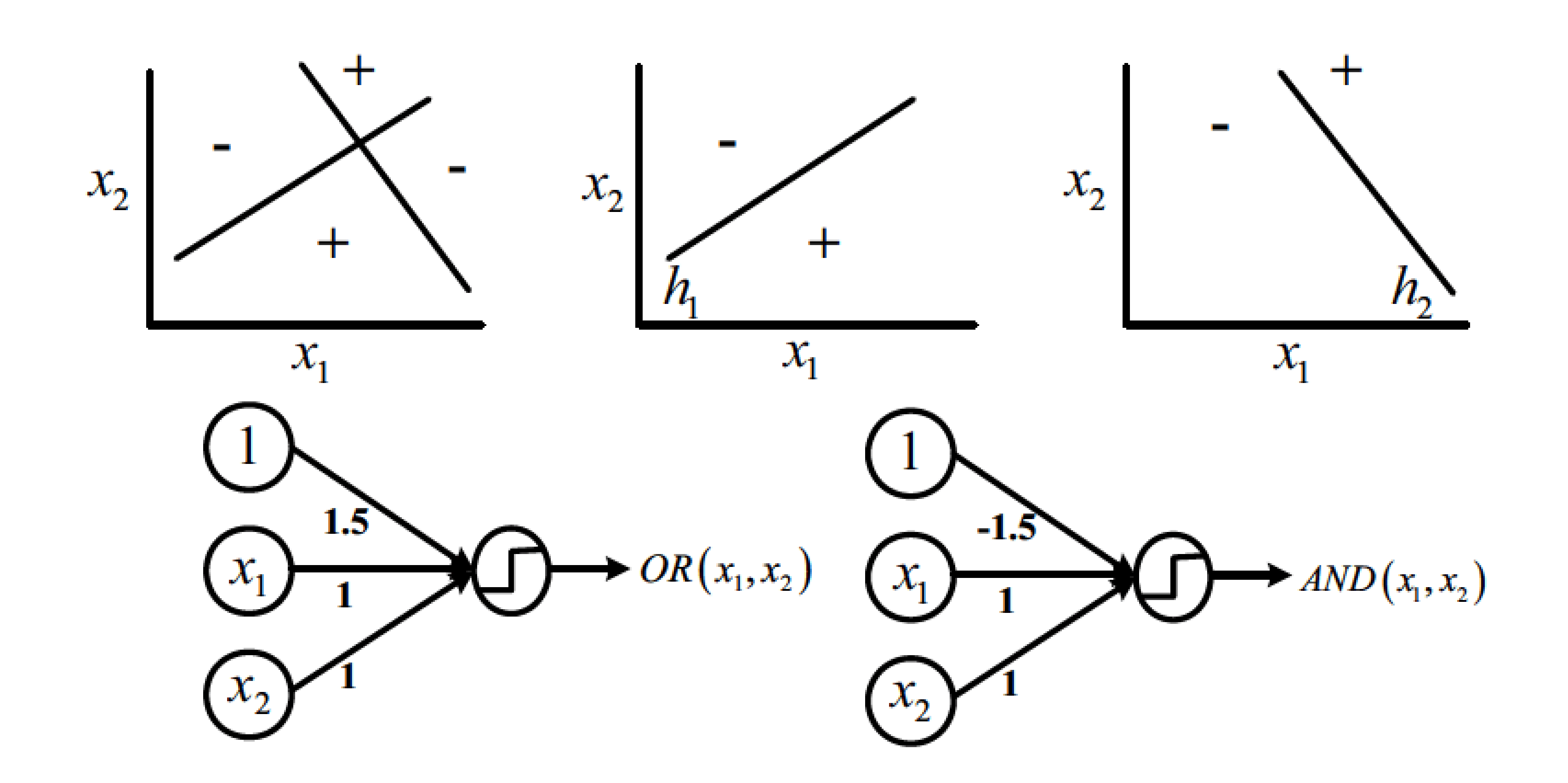


图 4.33　感知器

一个感知器相当于一个线性分类器，一层神经网络有多个隐藏节点时，为多个感知器的组合，就是多个线性分类器组合形成非线性分类器，如图 4.34 所示。

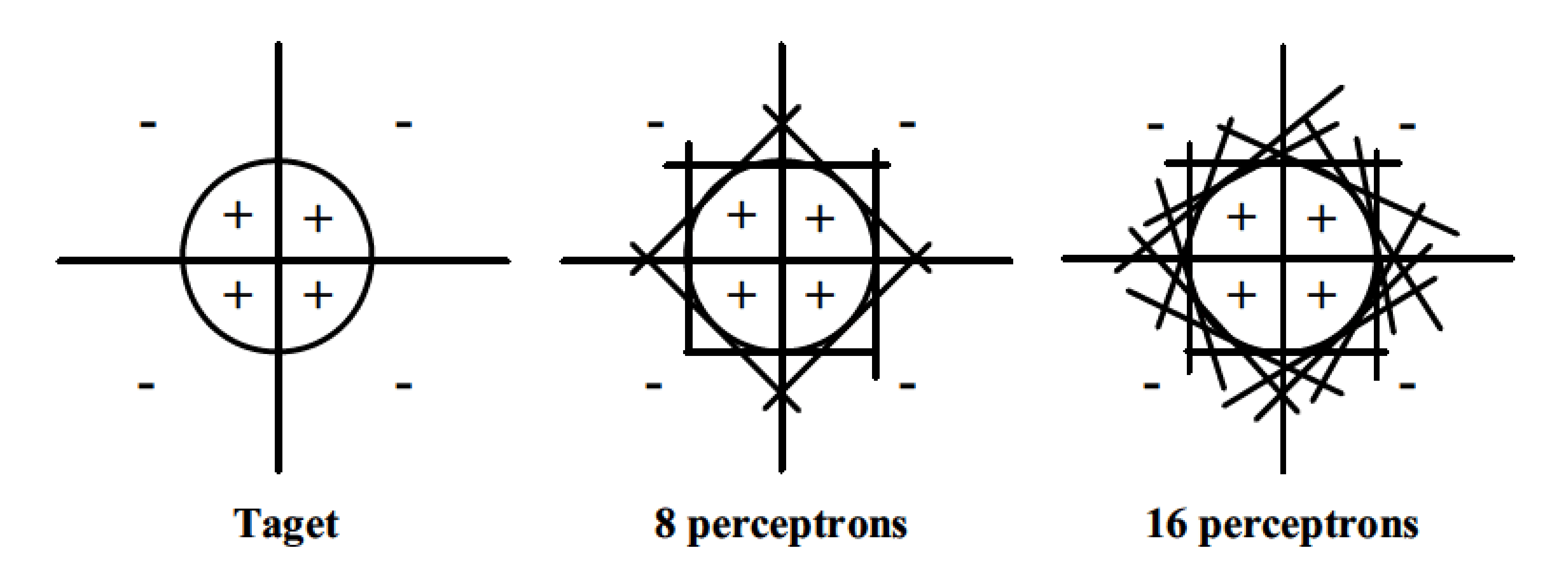


图 4.34　多感知器

多层的感知器组合起来，拟合能力虽然强大，但是求出准确拟合参数的算法不是太好，容易陷入局部最小，而且 BP 算法很容易陷入局部最小。

局部最小的情况，如图 4.35 所示，网络的权重被随机初始化后，求得梯度，然后用梯度更新参数，如果初始化的参数的点选择得不恰当，则梯度为 0 的点可能是一个使得代价 J 局部最小的点，而不是全局最小，自然得到的网络权重也不是最好的。因为网络规模大容易导致过拟合，深度学习提了一系列的 trick 改善这些问题。比如用贪心预训练来改进初始化参数，相当于找到一个好的初始点，在正负阶段里主动修改 J 的地形，最后再结合标签用传统的 BP 算法继续寻找全局最小，这个 BP 算法的作用在深度学习里叫权重微调，BP 不是唯一的权重微调算法，各种微调的宗旨只有一个：求取目标函数的梯度，更新参数。深度学习利用稀疏和 dropout 来阻止过拟合。

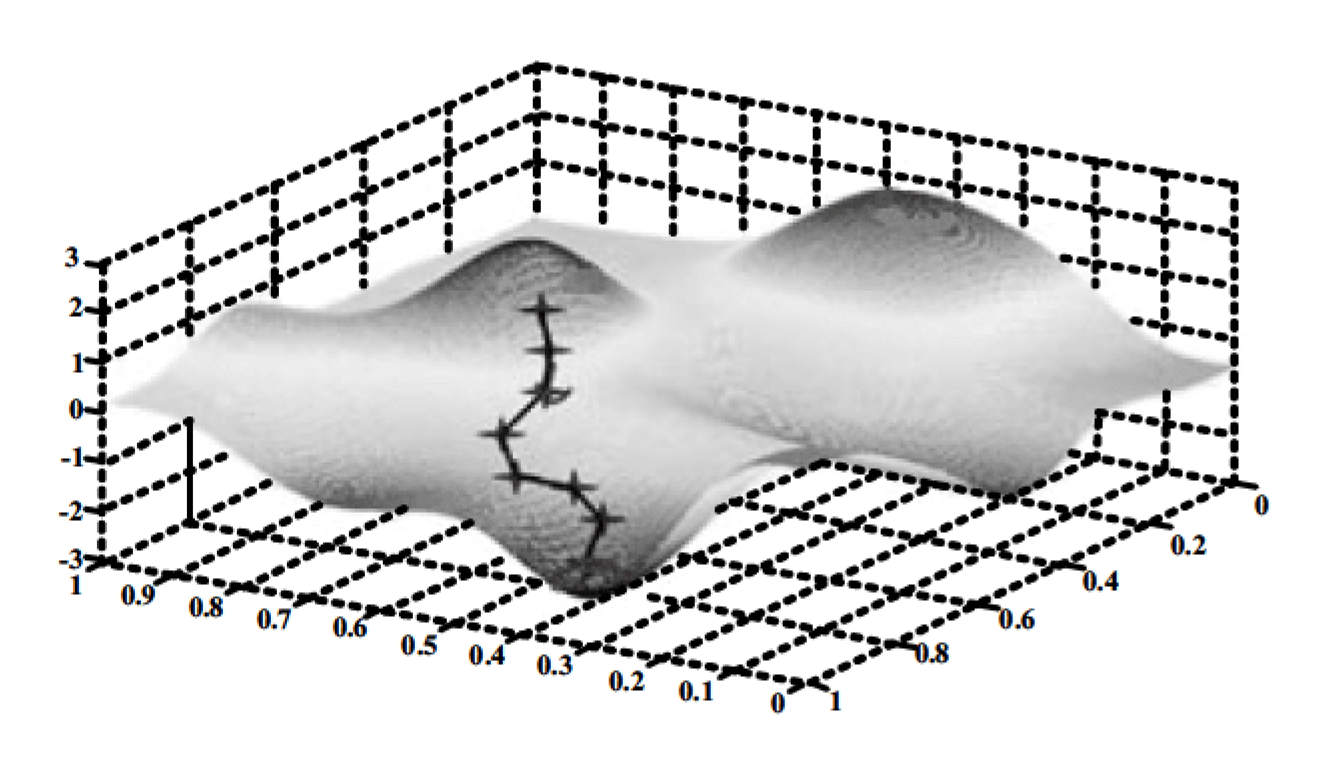


图 4.35　局部最小

4.8　卷积神经网络（CNN）

卷积神经网络是一种特殊的深层的神经网络模型，其特殊性体现在两个方面，一方面神经元间的连接是非全连接的，另一方面同一层中某些神经元之间的连接的权重是共享的（即相同的）。卷积神经网络具有非全连接和权值共享的网络结构，更类似于生物神经网络，可以降低了网络模型的复杂度，减少了权值的数量。

卷积网络最初受视觉神经机制的启发，设计为用于进行二维形状识别，对平移、缩放、倾斜等变形具有较高的不变性。1962 年 Hubel 和 Wiesel 在对猫的视觉皮层细胞进行研究后，提出感受野（receptive field）的概念，1984 年日本科学家 Fukushima 基于感受野概念提出神经认知机（neocognitron）模型，该模型将一个视觉模式分解为若干特征子模式，然后以分层递阶式相连的特征平面进行处理，试图将视觉系统模型化，并且利用位移恒定能力从激励模式中学习，使其能够在即使物体有位移或轻微变形时，可识别这些模式的变化形式。神经认知机被看作是第一个实现了的卷积神经网络，也是感受野概念在人工神经网络领域的首次应用。Fukushima 将神经认知机主要用于手写数字识别，其他科研工作者发展出多种卷积神经网络形式，广泛应用于邮政编码识别、车牌识别和人脸识别等方面。

1. CNN 的结构

卷积网络是在有监督方式下学会的，网络结构主要有稀疏连接和权值共享两个特点，包括如下形式的约束。

- 特征提取

每一个神经元从上一层的局部接受域得到突触输入，迫使它提取局部特征。一旦一个特征被提取出来，只要它相对于其他特征的位置被近似地保留下来，它的精确位置就

没有那么重要了。

❑ 特征映射

网络的每一个计算层都是由多个特征映射组成的，每个特征映射都是平面形式的。平面中单独的神经元在约束下共享相同的突触权值集，这种结构形式具有平移不变性、自由参数数量的缩减。

❑ 子抽样

每个卷积层都有与之相连的计算层，用于实现局部平均和子抽样，从而降低特征映射的分辨率，增强对平移和其他变形的适应性。

（1）稀疏连接（Sparse Connectivity）

卷积网络通过在相邻两层之间强制使用局部连接模式来利用图像的空间局部特性，在第 m 层的隐层单元只与第 m-1 层的输入单元的局部区域有连接，第 m-1 层的这些局部区域被称为空间连续的接受域。

设第 m-1 层为视网膜输入层，第 m 层的接受域的宽度为 3，也就是说该层的每个单元与且仅与输入层的 3 个相邻的神经元相连，第 m 层与第 m+1 层具有类似的链接规则，如图 4.36 所示。

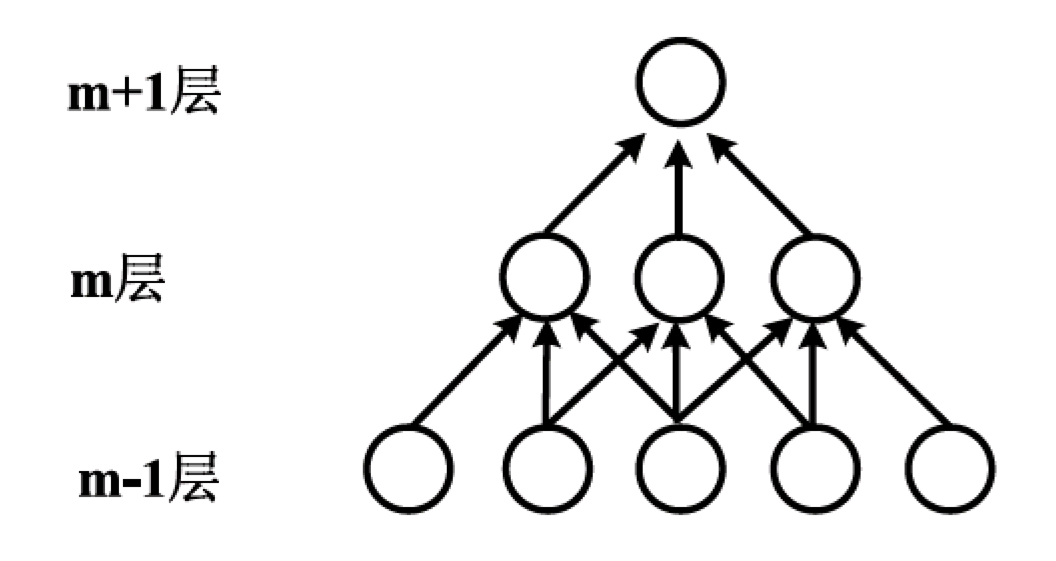


图 4.36 稀疏连接

可以看到 m+1 层的神经元相对于第 m 层的接受域的宽度也为 3，但相对于输入层的接受域为 5，这种结构将学习到的过滤器（对应于输入信号中被最大激活的单元）限制在局部空间模式，因为每个单元对它接受域外的 variation 不做反应。多个这样的层堆叠起来后，会使得过滤器（不再是线性的）逐渐成为全局的（也就是覆盖到更大的视觉区域）。如图 4.36 中第 m+1 层的神经元可以对宽度为 5 的输入进行一个非线性的特征编码。

（2）权值共享（Shared Weights）

在卷积网络中，每个稀疏过滤器 h_i 通过共享权值，覆盖整个可视域，这些共享权值的单元构成一个特征映射，如图 4.37 所示。

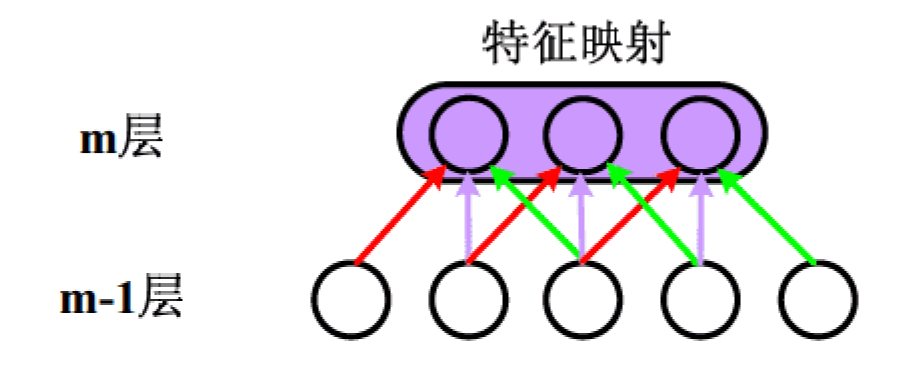


图 4.37　权值共享

在图 4.37 中，有 3 个隐含层单元，属于同一个特征映射。同种颜色的链接的权值是相同的，仍然可以使用梯度下降方法来学习这些权值，只需要对原始算法做一些小改动，共享权值的梯度是所有共享参数的梯度的总和。为什么要共享权值呢？一方面，重复单元能够对特征进行识别，而不考虑它在可视域中的位置。另一方面，权值共享能更有效地进行特征抽取，极大地减少需要学习的自由变量的个数。通过控制模型的规模，卷积网络对视觉问题可以具有很好的泛化能力。

（3）The Full Model

卷积神经网络具有多层结构，每一层由多个二维平面组成，每个平面又由多个独立神经元组成。网络中包含简单元和复杂元，分别记为 S-元和 C-元。S-元聚合在一起组成 S-面，S-面聚合在一起组成 S-层，用 Us 表示。类似地有 C-元、C-面和 C-层(Us)。卷积神经网络的输入只包含一层，可直接接入二维图像，中间级由 S-层与 C-层串接而成，卷积神经网络模型的互联结构实现特征提取。

Us 为特征提取层，内含神经元的输入为前一层的局部感受野，并提取该局部的特征，一旦该局部特征被提取后，它与其他特征之间的位置关系也随之确定下来；Uc 是特征映射层，多个特征映射组成网络的每个计算层，每个特征映射为一个平面，平面上所有神经元的权值相等。特征映射结构采用影响函数核小的 Sigmoid 函数作为激活函数，使特征映射具有位移不变性。一个映射面上的神经元共享权值，从而可以减少网络自由参数的个数，降低网络参数选择的复杂度。卷积神经网络中的每一个特征提取层（S-层）都紧跟着用来求局部平均与二次提取的计算层（C-层），这种特有的二次特征提取结构使网络对输入样本有较高的畸变容忍能力。

如图 4.38 所示，卷积网络的实现流程如下。

输入层由 32×32 个感知节点组成，接收原始图像数据。

计算流程在卷积和抽样之间交替进行：第一隐含层 C1 由 8 个特征映射组成，进行卷积运算，每个特征映射由 28×28 个神经元组成，每个神经元指定一个 5×5 的接受域；第二隐含层 S2 实现子抽样和局部平均，同样由 8 个特征映射组成，每个特征映射由 14×14 个神经元组成。每个神经元具有一个 2×2 的接受域、一个可训练系数、一个可训练偏置

和一个 Sigmoid 激活函数。可训练系数和偏置控制神经元的操作点。第三隐含层 C3 进行第二次卷积，由 20 个特征映射组成，每个特征映射由 10×10 个神经元组成。该隐含层中的每个神经元可能具有和下一个隐含层几个特征映射相连的突触连接，以与第一个卷积层相似的方式操作。第四个隐含层 S4 进行第二次子抽样和局部平均计算，由 20 个特征映射组成，每个特征映射由 5×5 个神经元组成，以与第一次抽样相似的方式操作。第五个隐含层 C5 实现卷积的最后阶段，由 120 个神经元组成，每个神经元指定一个 5×5 的接受域。

最后是全连接层，得到输出向量。

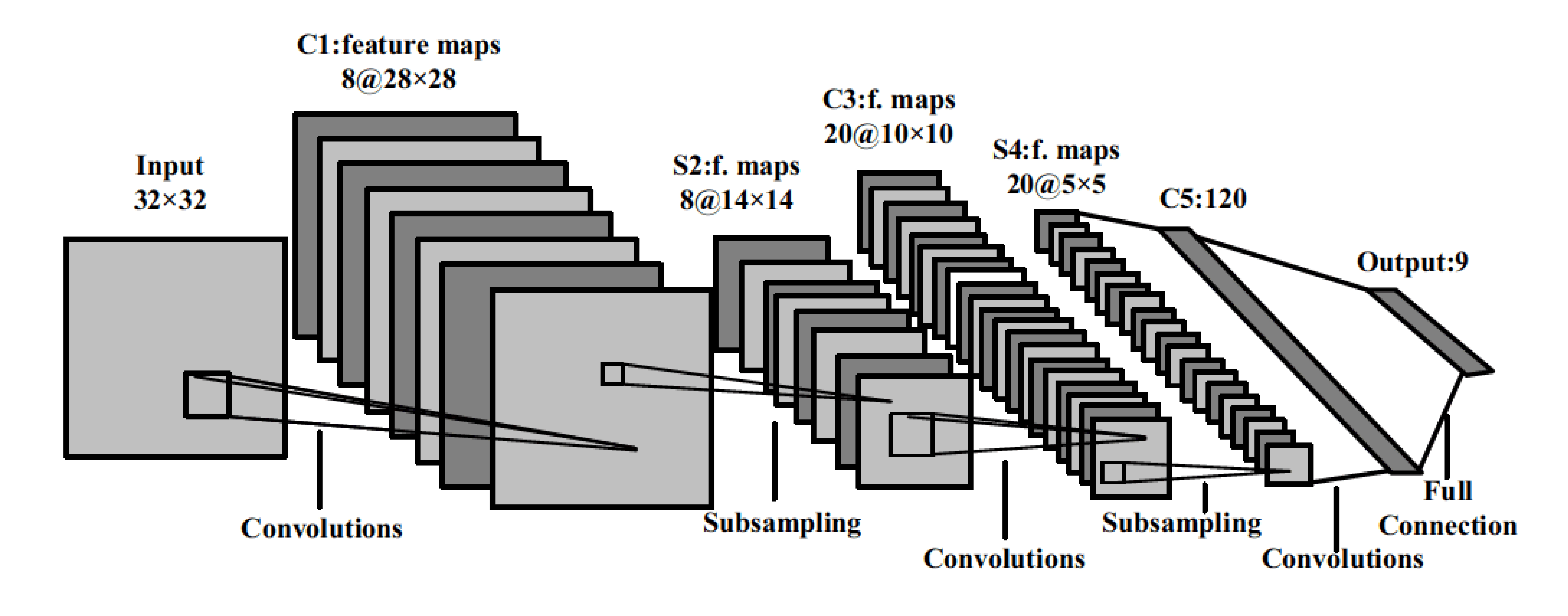


图 4.38 卷积网络实例

相继的计算层在卷积和抽样之间的连续交替，得到一个双尖塔的效果，也就是在每个卷积或抽样层，随着空间分辨率的下降，与相应的前一层相比特征映射的数量增加。卷积之后进行子抽样的思想产生于动物视觉系统中简单细胞后面跟着复杂细胞的启发。

图 4.38 中所示的多层感知器包含近似 100,000 个突触连接，但只有大约 2600 个自由参数。自由参数在数量上显著地减少，是通过权值共享获得的，学习机器的能力因而下降，提高了泛化能力。而且对自由参数的调整通过反向传播学习的随机形式来实现。另一个显著的特点是使用权值共享使以并行形式实现卷积网络变得可能。

2. CNN 的学习

卷积网络可以简化为图 4.39 所示的模型。

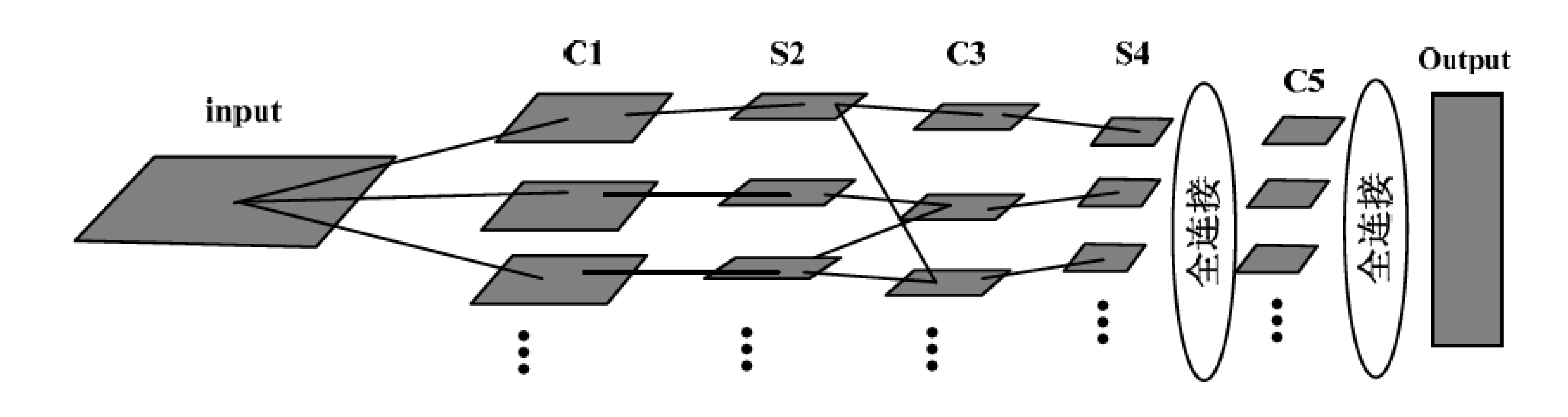


图 4.39　卷积网络

其中，input 到 C1、S4 到 C5、C5 到 output 是全连接，C1 到 S2、C3 到 S4 是一一对应的连接，为了消除网络对称性，S2 到 C3 去掉了一部分连接，可以让特征映射更具多样性。C5 卷积核的尺寸要和 S4 的输出相同，才能保证输出是一维向量。

（1）卷积层的学习

卷积层的典型结构如图 4.40 所示。

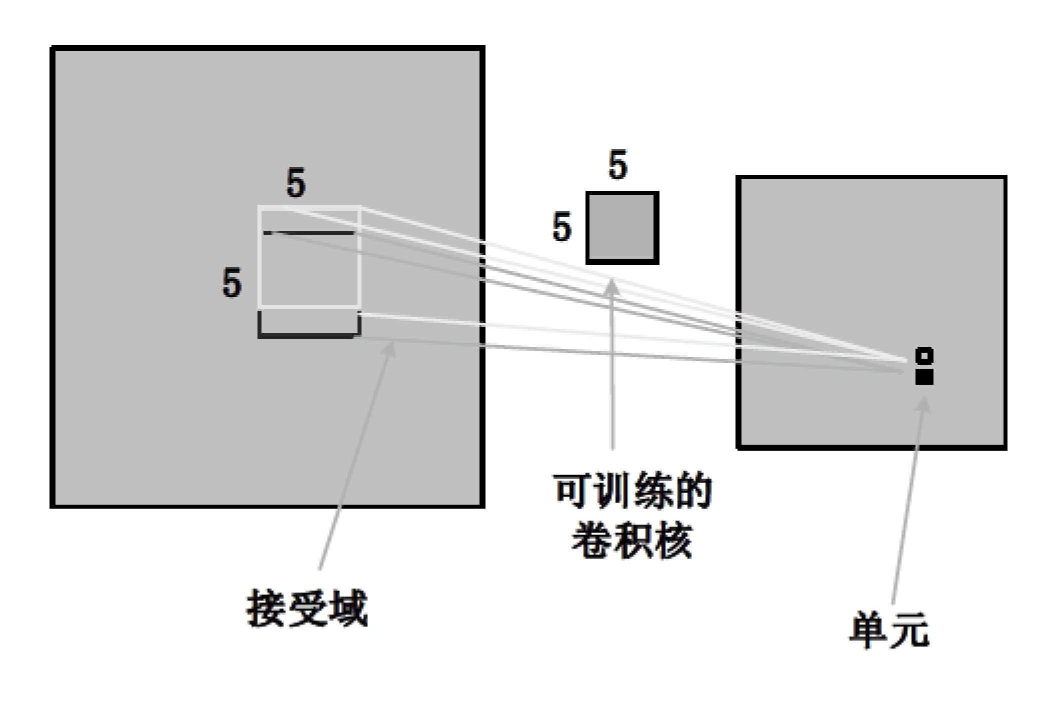


图 4.40　卷积层

卷积层的前馈运算为：

$$\text{卷积层的输出} = \text{Sigmoid}(\text{Sum}(\text{卷积}) + \text{偏移量})$$

其中卷积核和偏移量都是可训练的，其核心代码为：

```
fprop(input,output)
{
   //取得卷积核的个数
   int n=kernel.GetNum(0);
   for (int i=0;i<n;i++) {
     //第 i 个卷积核对应输入层第 a 个、输出层第 b 个特征映射
     //从输入层第 a 个特征映射到输出层第 b 个特征映射的一个链接
     int a=table[i][0], b=table[i][1];
     //用第 i 个卷积核和输入层第 a 个特征映射做卷积
     convolution = Conv(input[a],kernel[i]);
```

```
        //把卷积结果求和
        sum[b] +=convolution;
      }
      for (i=0;i<(int)bias.size();i++) {
        //加上偏移量
        sum[i] += bias[i];
      }
      //调用 Sigmoid 函数
      output = Sigmoid(sum);
    }
```

其中，input 矩阵的维数为 $n_1 \times n_2 \times n_3$，$n_1$ 是输入层特征映射的个数，n_2 表示输入层特征映射的宽度，n_3 为输入层特征映射的高度。Output、sum、convolution、bias 都是 $n_1 \times (n_2 - k_w + 1) \times (n_3 - k_h + 1)$的矩阵，$k_w$、$k_h$ 分别为卷积核的宽度、高度，一般选用 5×5。kernel 是卷积核矩阵。table 是连接表，其元素的意义为：如果第 a 个输入和第 b 个输出之间有连接，table 里 [a,b]元素取 1，否则取 0，而且每个连接都对应一个卷积核。

卷积层的反馈运算的核心代码为：

```
    ConvolutionLayer::bprop(input,output,in_dx,out_dx)
    {
      //梯度通过 DSigmoid 反传
      sum_dx = DSigmoid(out_dx);
      //计算 bias 的梯度
      for (i=0;i<bias.size();i++)  {
         bias_dx[i] = sum_dx[i];
      }
      //取得卷积核的个数
      int n=kernel.GetDim(0);
      for (int i=0;i<n;i++)
      {
         int a=table[i][0],b=table[i][1];
         //用第 i 个卷积核和第 b 个输出层反向卷积（即输出层的点乘
         //卷积模板返回给输入层），并把结果累加到第 a 个输入层
         input_dx[a] += DConv(sum_dx[b],kernel[i]);
         //用同样的方法计算卷积模板的梯度
         kernel_dx[i] += DConv(sum_dx[b],input[a]);
      }
    }
```

其中 in_dx、out_dx 的结构和 input、output 相同，代表相应点的梯度。

（2）子采样层的学习

子采样层的典型结构如图 4.41 所示。

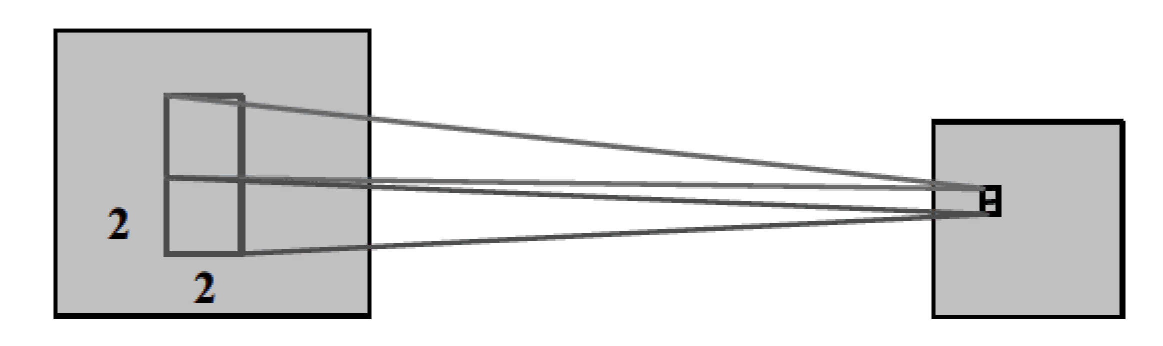


图 4.41　子采样层

子采样层的输出计算为：

$$输出 = \mathrm{Sigmoid}(\ 采样\times权重\ +\ 偏移量)$$

其核心代码为：

```
SubSamplingLayer::fprop(input,output)
{
   int n1= input.GetDim(0);
   int n2= input.GetDim(1);
   int n3= input.GetDim(2);
   for (int i=0;i<n1;i++) {
      for (int j=0;j<n2;j++) {
          for (int k=0;k<n3;k++) {
             //coeff 是可训练的权重，sw、sh 是采样窗口尺寸
             sub[i][j/sw][k/sh] += input[i][j][k]*coeff[i];
          }
       }
   }
   for (i=0;i<n1;i++) {
      //加上偏移量
      sum[i] = sub[i] + bias[i];
   }
   output = Sigmoid(sum);
}
```

子采样层的反馈运算的核心代码为：

```
SubSamplingLayer::bprop(input,output,in_dx,out_dx)
{
   //梯度通过 DSigmoid 反传
   sum_dx = DSigmoid(out_dx);
```

```
    //计算 bias 和 coeff 的梯度
    for (i=0;i<n1;i++) {
       coeff_dx[i] = 0;
       bias_dx[i] = 0;
       for (j=0;j<n2/sw;j++)
           for (k=0;k<n3/sh;k++) {
               coeff_dx[i] += sub[j][k]*sum_dx[i][j][k];
               bias_dx[i] += sum_dx[i][j][k]);
           }
    }
    for (i=0;i<n1;i++) {
       for (j=0;j<n2;j++)
           for (k=0;k<n3;k++) {
               in_dx[i][j][k] = coeff[i]*sum_dx[i][j/sw][k/sh];
           }
    }
}
```

全连接层的学习与传统的神经网络的学习方法类似，也使用 BP 算法，此处不再赘述。

4.9 AdaBoost 方法

AdaBoost（Adaptive Boosting）方法由美国加利福尼亚大学（University of California, San Diego）的 Yoav Freund 和美国普林斯顿大学（Princeton University）的 Robert E. Schapire 于 1995 年在 ECCLT 会议上提出，该方法深入挖掘弱分类器能力，不需要预先知道弱分类器的误差，得到的强分类器的分类精度依赖于所有弱分类器的分类精度。对应论文为 *A decision-theoretic generalization of on-line learning and an application to boosting*。

1. 基本原理

AdaBoost 方法是一种迭代过程，通过不断训练弱分类器，构成强分类器，从而提高数据分类能力。AdaBoost 方法的训练过程中，初始阶段每个样本具有相同的对应权重，在此样本分布下训练出一个弱分类器。然后针对分类错误的样本，加大其对应的权重；针对分类正确的样本则降低其权重，使前一步被分错的样本得到突显，获得新的样本分布。在新的样本分布下，再次对样本进行训练，又得到一个弱分类器。依次类推，经过 T 次循环，得到 T 个弱分类器，将这 T 个弱分类器按一定的权重组合，得到最终想要的

强分类器。

AdaBoost 方法是经过调整的 Boosting 算法，能够对弱学习得到的弱分类器的错误进行适应性调整。相对于 Boosting 方法，AdaBoost 方法使用加权后选取的训练数据代替随机选取的训练样本，训练的关键是针对比较难分的训练数据样本；在联合弱分类器时，使用加权投票机制代替平均投票机制。通过上述处理，分类效果好的弱分类器将获得较大的权重，而分类效果差的分类器则权重较小。

2. AdaBoost 的实现

给定训练集 $(x_1, y_1),\cdots,(x_N, y_N)$，其中 $y_i \in \{1,-1\}$，表示 x_i 的正确类别标签，$i=1,\cdots,N$。在训练集上样本的初始分布为：

$$D_1(i)=\frac{1}{N}$$

对 $t=1,\cdots,T$，计算弱分类器：

$$h_t : X \to \{-1,1\}$$

该弱分类器在分布 D_t 上的误差为：

$$\varepsilon_t = \mathbb{P}_{D_t}\left(h_t(x_i) \neq y_i\right)$$

计算该弱分类器的权重：

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

更新训练样本的分布：

$$D_{t+1}(i)=\frac{D_t(i)\exp\left(-\alpha_t y_i h_t(x_i)\right)}{Z_t}$$

其中 Z_t 为归一化常数。

最后的强分类器为：

$$H_{final}(x)=sign\left(\sum_{t=1}^{T}\alpha_t h_t(x)\right)$$

3. AdaBoost 的权值

对于每次迭代要把错分点的权值变大，AdaBoost 的表达式为：

$$H_{final}(x)=sign\left(\sum_{t=1}^{T}\alpha_t h_t(x)\right)$$

其中的α表示权值，是关于误差的表达式：

$$\alpha_t=\frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

提高错分点的权值，当下一次分类器再次分错这些点之后，会提高整体的错误率，导致α变得很小，最终这个分类器在整个混合分类器的权值变低。这样，让优秀分类器的权值更高，一般分类器的权值更低。

4. AdaBoost 的流程

AdaBoost 方法的实现流程如下。

步骤01　给定训练样本集 S，其中 X 和 Y 分别为正样本和负样本；T 为训练的最大循环次数；

步骤02　初始化样本权重为 $1/n$，即为训练样本的初始概率分布；

步骤03　循环迭代多次：

- 更新样本权重和分布;
- 寻找当前分布下的最优弱分类器;
- 计算弱分类器误差率;
- 选取合适阈值，使误差最小。

步骤04　聚合多次训练的弱分类器。

经 T 次循环后，得到 T 个弱分类器，按更新的权重叠加，最终得到强分类器。

5. AdaBoost 的伪代码

AdaBoost 方法的实现伪代码如下。

已知：$(x_1,y_1),\cdots,(x_m,y_m)$，其中 $x_i\in X, y_i\in Y=\{-1,+1\}$

初始化：$D_1(i)=1/m$

For　$t=1,\cdots,T$:

利用分布 D_t 训练弱分类器。

得到弱分类器 $h_t: X \rightarrow \{-1,+1\}$，对应的误差为

$$\varepsilon_t = \Pr_{i \sim D_t}\left[h_t(x_i) \neq y_i\right]$$

选取 $\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$。

更新：

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & if\ h_t(x_i) = y_i \\ e^{\alpha_t} & if\ h_t(x_i) \neq y_i \end{cases}$$
$$= \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

其中，Z_t 为归一化因子。

输出最终的分类器：

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

6. 算法特点

AdaBoost 方法是一种高精度的分类器，可以使用各种方法构建子分类器，而且弱分类器构造简单，不需要进行特征筛选，也不用担心过度拟合。

AdaBoost 方法可用于二分类或多分类的应用场景，可用于特征选择，只需要增加新的分类器，不需要变动原有分类器。是一种实现和应用很简单的算法，通过组合弱分类器得到强分类器，分类错误率上界随着训练的增加而稳定下降，不会过拟合，适合于各种分类场景。

在 AdaBoost 方法训练过程中，每次迭代都会对分类错误的样本进行加权，当出现多次分类错误以后，它们的权重过大，进而影响误差的计算和分类器的挑选，使分类器的精度下降，即典型退化问题。这些样本往往是靠近分类边界的样本，称为临界样本。临界样本使得训练的退化问题加剧，但也是提升分类器精度的必需品。

在某些应用（如车牌检测处理）中，现实中车牌的数目要远远小于非车牌数，负样本的范围非常广，样本集往往无法精确表示，正负样本的数量差距很大，分类器会关注大容量样本，导致分类器不能较好地完成区分小类样本的目的。数据不平衡问题是 AdaBoost 方法的一个典型难题。

4.10　模拟退火方法

模拟退火方法（Simulated Annealing Algorithm，SAA）是 IBM 的 S.Kirkpatrick 等人于 1983 年在研究组合优化的基础上，根据迭代改进思想提出的。它是一种通用概率算法，用来在固定时间内寻求在一个大的搜寻空间内找到最优解。

在某个定义域 S 内，求某个函数 $f(x)$的最小值，形式化为 Min $f(x)$，x 属于 S。在搜索极值过程中，如果过早结束，就会陷入局部最优情况，为了跳出局部最优，引入一个接受概率 P 和参数 T。在当前解的邻域内选择一点，如果比当前解好，则总是接受它；如果没有当前解好，则以接受概率接受它。接受概率中的 T 随着时间从大到小变化（冷却温度），一开始 T 值很大，近似于随机搜索，随机选择当前解；后来 T 很小，近似于普通搜索法，选择最优作为当前解。

1. 爬山方法

爬山方法是一种简单的贪心搜索算法，每次从当前解的临近解空间中选择一个最优解作为当前解，直到达到一个局部最优解。

爬山方法的实现很简单，主要缺点是会陷入局部最优解，而不一定能搜索到全局最优解。如图 4.42 所示，假设 C 点为当前解，爬山方法搜索到 A 点这个局部最优解就会停止搜索，因为在 A 点无论向哪个方向小幅度移动都不能得到更优的解。

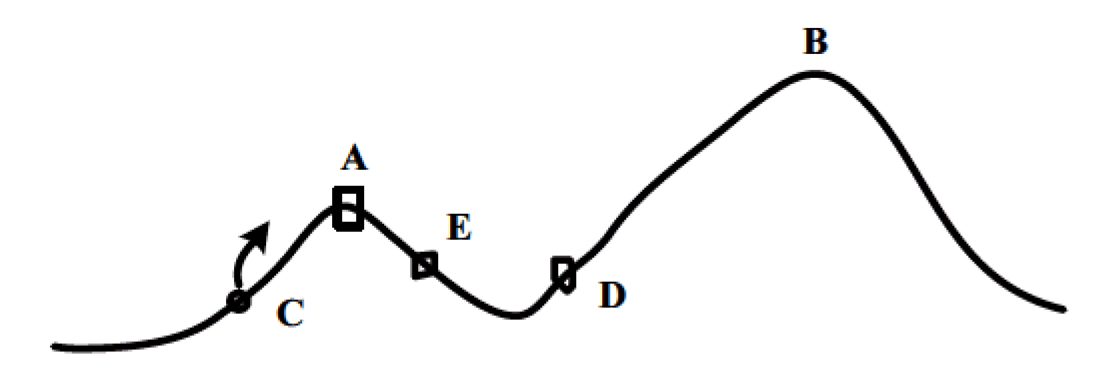


图 4.42　爬山方法示意图

2. 模拟退火思想

爬山方法是完完全全的贪心法，每次都鼠目寸光地选择一个当前最优解，因此只能搜索到局部最优值。模拟退火方法也是一种贪心算法，但是搜索过程引入随机因素。以一定的概率来接受一个比当前解要差的解，有可能会跳出这个局部的最优解，达到全局最优解。以图 4.42 为例，模拟退火方法在搜索到局部最优解 A 后，会以一定的概率接受到 E 的移动。经过几次不是局部最优的移动后会到达 D 点，于是就跳出局部最大值 A。

模拟退火思想描述为：

- 若 $J(Y(i+1))>=J(Y(i))$，即移动后得到更优解，则总是接受该移动；

- ❑ 若 $J(\mathrm{Y}(i+1))< J(\mathrm{Y}(i))$，即移动后的解比当前解要差，则以一定的概率接受移动，而且这个概率随着时间推移逐渐降低，趋向稳定。

这里对一定的概率的计算参考了金属冶炼的退火过程，根据热力学原理，在温度为 T 时，出现能量差为 dE 的降温的概率为 P(dE)，表示为：

$$P\left(dE\right)=\exp\left(dE/\left(kT\right)\right)$$

其中，k 是常数，exp 表示自然指数，且 dE<0。温度越高，出现一次能量差为 dE 的降温概率越大；温度越低，出现降温概率越小。由于 dE<0，dE/kT <0，所以 P(dE)的函数取值范围是(0,1)。

随着温度 T 的降低，P(dE)会逐渐降低。将一次向较差解的移动看做一次温度跳变过程，以概率 P(dE)接受这样的移动。

3. 模拟退火方法

模拟退火方法所得解依据概率收敛到全局最优解。首先建立数学模型，包括要确定解空间，确立目标函数和初始解；然后在产生新解时要符合某种接受机制；最后由接受准则使新解更优或是恶化。

数学模型由解空间、目标函数和初始解 3 部分组成。

（1）解空间

当所有可能解均为可行解时，解空间为可能解的集合；针对不可行解，一种情况是限定解空间为所有可行解集，另一种方法为允许包含不可行解，但在目标函数中通过罚函数排除不可行解。

（2）目标函数

目标函数是从解空间到某个数集的映射，表示为对优化目标的量化描述，应正确体现问题的整体优化要求，并且需便于计算，当解空间包含不可行解时还应包括罚函数项。

（3）初始解

算法迭代的起点。模拟退火方法是一种最终解，不强烈依赖于初始数据的健壮算法，因此可随机选取初始解。

新解的产生和接受流程包括 4 个步骤：首先，按某种随机方法由当前解产生一个新解，通常利用简单变换产生，如部分元素的置换、互换或反演等，将可能产生的新解作为当前解的邻域；接着，由变换的改变部分计算新解伴随的目标函数差；然后，根据接受原则，即新解是否更优或恶化但满足 Metropolis 准则，判断是否接受新解，并且还需判断其

解的可行性；最后，满足接受准则时进行当前解和目标函数值的迭代，否则舍弃新解。

4. 模拟退火方法的伪代码

模拟退火方法的算法伪代码为：

```
/* J(y)：在状态 y 时的评价函数值
* Y(i)：表示当前状态
* Y(i+1)：表示新的状态
* r：用于控制降温的快慢
* T：系统的温度，系统初始应该要处于一个高温的状态
* T_min：温度的下限，若温度 T 达到 T_min，则停止搜索
*/
While ( T > T_min )
{
    dE = J( Y(i+1) ) - J( Y(i) );
    if ( dE >= 0 )  //表达移动后得到更优解，则总是接受移动
    Y(i+1) = Y(i);  //接受从 Y(i)到 Y(i+1)的移动
    Else
    {
        //函数 dE/T 越大，则 exp(dE/T)也越大
        if ( exp( dE/T ) > random( 0 , 1 ) )
        Y(i+1) = Y(i);  //接受从 Y(i)到 Y(i+1)的移动
    }
    T = r * T;  //降温退火，0<r<1。r 越大，降温越慢
    /* 若 r 过大，则搜索到全局最优解的可能性会较高，但搜索过程较长。若 r 过小，则搜索
过程会很快，但可能会达到局部最优值 */
    i ++;
}
```

5. 算法特点

与局部搜索方法相比，模拟退火方法可在较短时间里求得更优近似解。允许任意选取初始解和随机数序列，能得出较优近似解，求解优化问题的前期工作量大大减少。在可能影响模拟退火方法实验性能的诸多因素中，问题规模 n 的影响最为显著，n 的增大导致搜索范围的绝对增大，会使 CPU 时间增加；而对于解空间而言，搜索范围又因 n 的增大而相对减小，引起解质量下降，但 SAA 的解和 CPU 时间均随 n 增大而趋于稳定，不受初始解和随机数序列的影响。该方法能应用于多种优化问题。

模拟退火方法是一种随机算法，并不一定能找到全局最优解，可以比较快地找到问题的近似最优解。如果参数设置得当，模拟退火方法搜索效率比穷举法要高。

模拟退火方法返回一个高质近似解的时间花费较多，当问题规模不可避免地增大时，难于承受的运行时间将使算法丧失可行性。选择适当的邻域结构和随机数序列可以提高解质并缩减运行时间，这需要大量试验。选择合理的冷却进度表可使算法的执行过程更有效。

模拟退火方法的控制参数对算法性能有一定的影响，没有一个适合各种问题的参数选择方法，只能依赖于具体问题进行确定。

4.11　遗传方法

遗传方法（Genetic Algorithm）起始于 20 世纪 60 年代，由美国密歇根大学的 John Holland 等提出，也称进化方法。它是受达尔文进化论的启发，借鉴生物进化过程而提出的一种启发式搜索方法。

1. 基本原理

遗传方法的重要概念如下。

- 染色体（Chromosome）：生物细胞中含有的一种微小的丝状化合物，是遗传物质的主要载体，由多个遗传因子（基因）组成。
- 遗传因子（gene）：DNA 长链结构中占有一定位置的基本遗传单位，也称基因，生物的基因根据物种的不同而多少不一。
- 个体（individual）：染色体带有特征的实体。
- 种群（population）：染色体带有特征的个体的集合。
- 进化（evolution）：生物在其延续生命的过程中，逐渐适应其生存环境，使品质不断得到改良。生物的进化是以种群形式进行的。
- 适应度（fitness）：度量某个物种对于生存环境的适应程度。
- 选择（selection）：指以一定的概率从种群中选择若干个体的操作。
- 变异（musation）：很小的概率产生的某些复制差错；亲代和子代之间，子代和子代的不同个体之间总有些差异，变异是随机发生的。
- 编码（coding）：DNA 中遗传信息在一个长链上按一定的模式排列，进行遗传编码。遗传编码可以看成是从表现型到遗传子型的映射。
- 解码（decoding）：从遗传子型到表现型的映射。

如图 4.43 所示，遗传方法是从代表问题可能潜在解集的一个种群开始的。该种群由经过基因编码的一定数目的个体组成。初代种群产生之后，按照适者生存、优胜劣汰的

原则，逐代进化产生出越来越好的近似种，即在每一代中，根据问题域中个体适应度大小挑选个体，并借助自然遗传学的遗传算子进行组合交叉和变异，产生出代表解的解集种群。这个过程将导致种群像自然进化一样，后生代种群比前代更加适应环境，末代种群中的最优个体经过解码可以作为问题近似最优解。

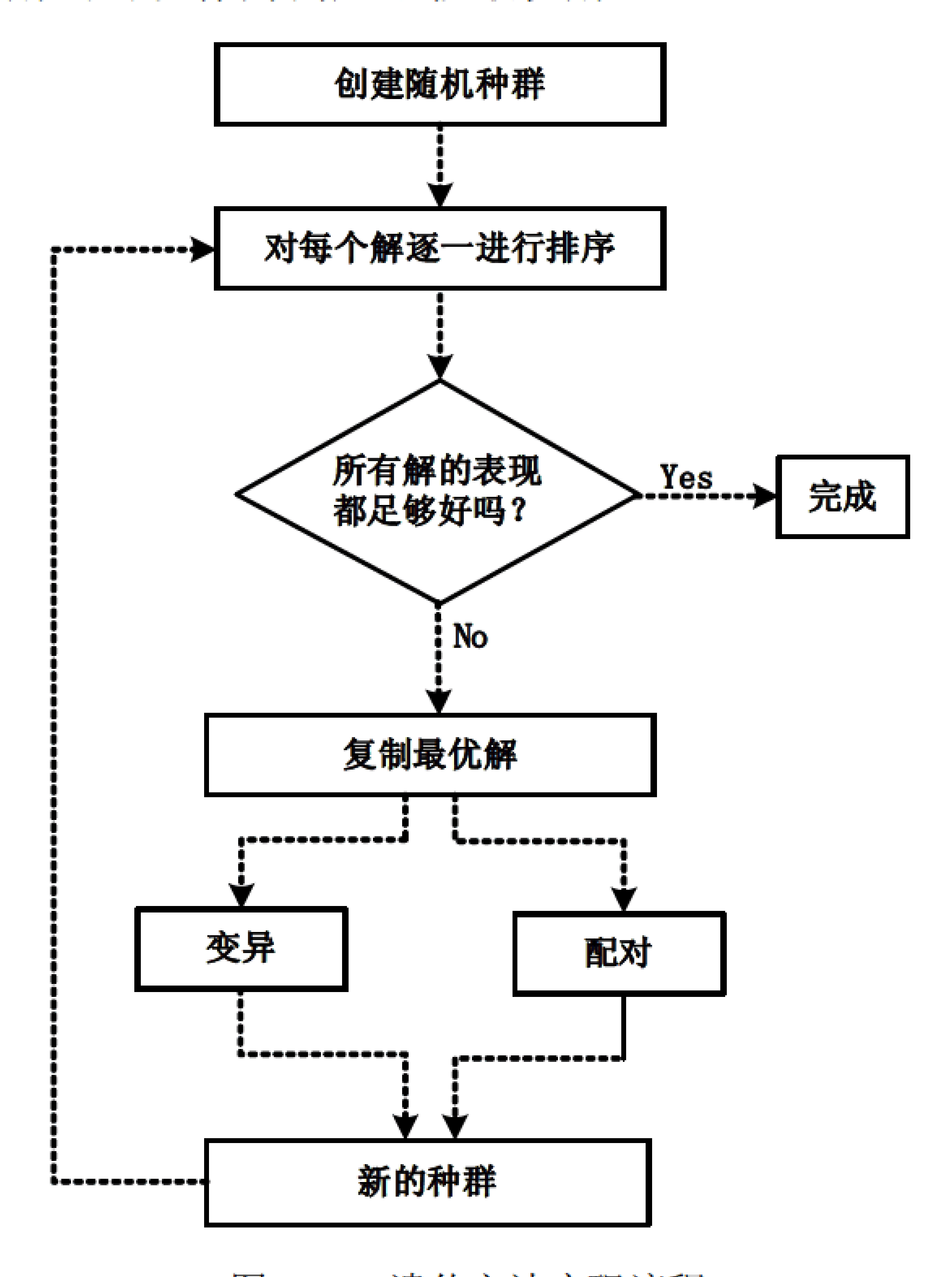


图 4.43　遗传方法实现流程

遗传方法模拟自然选择和遗传中发生的复制、交叉和变异等现象，从任一初始种群出发，通过随机选择、交叉、变异操作，逐步淘汰掉适应度函数值低的解，增加适应度函数值高的解，产生一群更适合环境的个体，使群体进入到搜索空间中越来越好的区域，这样一代一代地不断繁衍进化，最后收敛到一群最适合环境的个体，求得问题的最优解。

（1）编码方法的设计

Holland 提出的遗传方法利用二进制编码表现个体遗传基因类型，使用的编码符号集由二进制符号 0 和 1 组成，因此将遗传基因表现为二进制符号串。其优点是编解码操作简单，交叉、变异等遗传操作便于实现。缺点是不便于反映所求问题的特定性质，由于遗传方法的随机性而使其局部搜索能力较差，对于一些多维、高精度要求的连续函数优化，二进制编码进行离散化时存在映射误差，如果个体编码串较短，可能不满足精度

要求；如果个体编码较长，虽然能提高精度，但却导致搜索空间扩大，降低整体性能。为提高遗传方法的局部搜索能力，后来又产生了格雷码等编码方法。

遗传方法的进化过程建立在编码基础上，编码方法对搜索能力和种群多样性等性能影响很大，譬如二进制编码搜索能力强，而种群多样性弱，浮点编码正好相反。

根据具体问题确定待寻优的参数；对每个参数确定它的变化范围并用二进制表示：

$$a = a_{\min} + \frac{b}{2^m - 1}(a_{\max} - a_{\min}), a \in [a_{\min,} a_{\max}]$$

b 为 m 位二进制数，m 在满足精度要求下应尽量小。

将所有表示参数的二进制数串接起来组成二进制字符串，每一位只能取值 0 或 1，该字符串即为一串方法的操作对象。

浮点编码对个体 x_t^i 的第 p 个基因进行变异操作。

$$\psi_D(x_t^i p) = x_t^{i(p)} + N(0, \delta)$$

其中，N 为高斯噪声。

可见浮点数编码的变量可以任意小，并且只要变异量足够小，产生的新个体可以与父个体充分接近。而二进制编码的变异操作不能保证父个体与新个体充分接近，种群稳定性比浮点差。

（2）适应度函数的选取

遗传方法在进化搜索中主要以适应度函数为依据，利用种群中每个个体的适应度值进行搜索，基本不利用外部信息。因此适应度函数的选取至关重要，直接关系到收敛速度以及能否找到最优解。通过对目标函数值域的某种映射变换，可以成为适应度的尺度变换。

适应度函数需满足以下条件：单值、连续、非负、最大化、计算量小、通用性强。

直接以待求解的目标函数转化为适应度函数，若目标函数为最大化问题，则

$$Fit\big(f(x)\big) = f(x)$$

若目标函数为最小化问题，则

$$Fit\big(f(x)\big) = -f(x)$$

该适应度函数简单、直观，但有两个缺陷，其一是可能不满足概率非负要求；其二是某些待求解的函数在函数值分布上相差很大，平均适应度可能不利于体现种群的平均

性，影响方法性能。

若目标函数为最小问题，则

$$Fit(f(x))=\begin{cases} c_{\max}-f(x), & f(x)<c_{\max} \\ 0 & \text{其他} \end{cases}$$

式中 $C_{\max}$ 为 $f(x)$的最大值估计。反之，则

$$Fit(f(x))=\begin{cases} f(x)-c_{\min}, & f(x)>c_{\min} \\ 0 & \text{其他} \end{cases}$$

$C_{\min}$ 为 $f(x)$的最小值估计。

该方法是第一种方法的改进，称为界限构造法，但 $c_{\max}$ 与 $c_{\min}$ 的构造与选择困难。

若目标函数取为最小问题，则

$$Fit(f(x))=\frac{1}{1+c+f(x)} \qquad c\geq 0, c+f(x)\geq 0$$

若目标函数为最大问题，则

$$Fit(f(x))=\frac{1}{1+c-f(x)} \qquad c\geq 0, c-f(x)\geq 0$$

c 为目标函数界限的保守估计值。

（3）适应度函数的尺度变换

A. 线性变换法

$$f'=\alpha\times f+\beta$$

α和β的确定有多种方法，但是原适应度的平均值要等于标定后的适应度平均值，以保证适应度为平均值的个体在下一代的期望复制数为 1。

变换后的适应度最大值应等于原适应度平均值的最大倍数，以控制适应度最大的个体在下一代中的复制数。指定倍数 *Cmult* 可在 1.0~2.0 范围内。即根据上述条件可确定线性比例系数：

$$f'_{\max}=c_{mult}f_{avg}$$

$$\alpha = \frac{\left(c_{mult} - 1\right) f_{avg}}{f_{\max} - f_{avg}}$$

$$\beta = \frac{\left(f_{\max} - c_{mult} f_{avg}\right) f_{avg}}{f_{\max} - f_{avg}}$$

B. 幂函数变换法

$$f' = f^k$$

C. 指数变换法

$$f' = e^{-\alpha f}$$

这种变换法来源于模拟退火过程，系数 α 决定复制的强制性，其值越小，复制强度就越趋向于那些具有最大适应度的个体。

（4）选择过程

选择过程的第一步是计算适应度。在被选中集中的每个个体具有一个选择概率，这个选择概率取决于种群中个体的适应度。

按比例的适应度分配又称为选择的蒙特卡罗法，利用比例于各个个体适应度的概率决定其子孙的遗留可能性。

$$p_i = \frac{f_i}{\sum_{i=1}^{M} f_i}$$

其中，M 为个体总数目。

基于排序的适应度分配为种群按目标值进行排序。适应度仅仅取决于个体在种群中的序位，而不是实际的目标值。排序方法克服比例适应度计算的尺度问题，可通过引入种群均匀尺度来控制选择压力。排序方法表现出有效的鲁棒性。

设定 N 为种群大小，Pos 为个体在种群中的序位，SP 为选择压力，个体的适应度可以计算如下。

线性排序：

$$Fit(Pos) = 2 - SP + \frac{2(SP-1)(Pos-1)}{N-1} \qquad SP \in [1.0, 2.0]$$

非线性排序：

$$Fit(Pos) = \frac{NX^{Pos-1}}{\sum_{i=1}^{n} X^{i-1}}$$

其中 X 是下列多项式方程的根：

$$(SP-1)X^{N-1} + SPX^{N-2} + \ldots + SPX + SP = 0, \quad SP \in [1.0, N-2.0]$$

（5）基因交叉重组

基因交叉（重组）是把两个父个体的部分结构进行替换而重组产生下一代新的子个体的操作。基因重组是遗传方法获得新优良个体的重要手段。

根据编码表示的不同，可以有以下方法。

- 实值重组：离散重组、中间重组、线性重组、扩展线性重组。
- 二进制交叉：单点、多点、均匀、洗牌和缩小代理等交叉。

离散重组在个体之间交换变量值，考虑如下 3 个变量的个体：

父个体 1　　12　　25　　5
父个体 2　　123　　4　　34

子个体中每个变量可按等概率随机的挑选父个体，如：

子个体 1　　123　　4　　5
子个体 2　　12　　4　　5

中间重组只适用于实变量。

子个体 1=父个体 1+α（父个体 2-父个体 1）
$\alpha \in [-d, 1+d]$

对于中间重组 d=0，一般选择 d=0.25。

父 1　12　25　5
父 2　123　4　34
样本 1　0.5　1.1　-0.1
样本 2　0.1　0.8　0.5
子个体 1　12+0.5（123-12）=67.5
　　25+1.1（4-25）=1.9
　　5+（-0.1）（34-5）=2.1

线性重组与中间重组相似，也是对所有变量值有一个α值。

父 1　12　25　5
父 2　123　4　34
样本 1　0.5
样本 2　0.1
子 1：12+0.5(123-12)=67.5　25+0.5(4-25)=14.5　5+0.5(34-5)=19.5
子 2：12+0.1(132-12)=23.1　25+0.1(4-25)=22.9　5+0.1(34-5)=7.9

（6）变异过程

子个体变量以很小的概率或步长产生变异，该概率或步长与种群大小无关，与变量个数成反比。对于单峰函数而言，$1/n$ 是最好的平均选择，并且通过在开始时增加变异率，结束时减小变异率可以改善搜索速度。对于多峰函数而言，其变异率的自适应过程是很有益的选择。变异本身是一种局部随机搜索，与选择/重组算子结合在一起，使遗传方法具有局部的随机搜索能力，同时使遗传方法保持种群的多样性，以防止出现非成熟收敛。变异操作中变异率不能太大，否则就退化为纯随机搜索。

A. 实值变异

$$x' = x \pm 0.5L\Delta$$

$$\Delta = \sum_{i=0}^{m-1} \frac{a(i)}{2^i}$$

L 为变量的取值范围；$a(i)$以概率 $1/m$ 取值 1，以 $1-1/m$ 取值 0。

B. 二进制变异

此时变异即意味着翻转，其变异位是随机确定的。

变异前　0110011010
变异后　0111011010

另外还有换位、复制、插入、删除变异。

2. 实现方法

（1）遗传方法流程

如图 4.44 所示，遗传方法的实现流程如下。

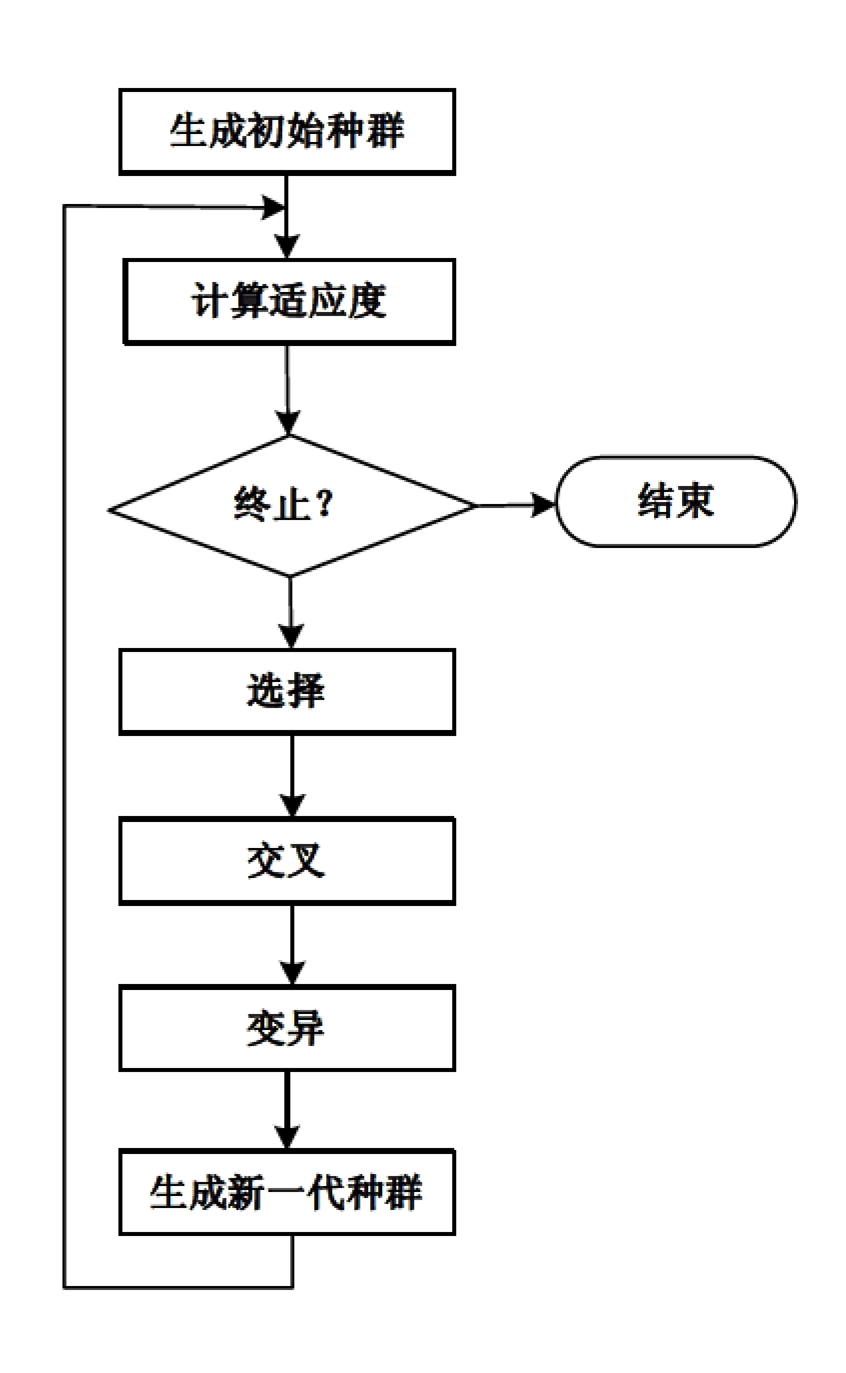


图 4.44 遗传方法的实现流程

步骤 01 编码

遗传方法在进行搜索之前，首先将解空间的数据表示成遗传空间的基本串结构数据类型，不同的组合就构成不同的点。

常用固定长度的二进制符号串。

步骤 02 初始种群的形成

随机生成 N 个初始串数据，每个串数据构成一个个体，由 N 个串数据构成一个群体。以该群体作为初始点开始迭代。

群体大小一般取 20~100。

步骤 03 适应度检测

适应度就是生物个体对环境的适应程度，适应度函数是对问题中的个体对象所设计的表征其优劣的测度。

根据具体问题计算 $P(t)$适应度。

步骤 04 选择

将选择算子作用于群体，把优化个体直接遗传到下一代。选择算子又称为再生算子，如按比例的适应度方法、基于排序的适应度方法、轮盘赌选择、随机遍历抽样、局部选择、截断选择。

常使用比例选择算子。

步骤 05　交叉

遗传方法中起核心作用的就是交叉算子，根据交叉率将种群中的两个个体随机地交换某些基因，产生新基因组合。

常使用单点交叉算子。

步骤 06　变异

变异算子首先对群中所有个体以事先设定的变异概率判断是否进行变异，然后对进行变异的个体随机选择变异位进行变异。

常使用基本位变异算子。

步骤 07　终止条件判断

群体 $P(t)$经过选择、交叉、变异运算之后得到下一代群体 $P(t+1)$。

若 $t<=T$，则 $t=t+1$，转到 *Step*3，否则以进化过程中所得到的具有最大适应度个体作为最优解输出，终止计算。

一般终止进化代数为 100~500。

（2）伪代码

基本遗传方法的伪代码如下。

```
Algorithm  GA(Pc,Pm,M,G,Tf)
Input:
Pc: 交叉发生的概率
Pm: 变异发生的概率
M: 种群规模
G: 终止进化的代数
Tf: 进化产生的任何个体的适应度函数超过 Tf，则终止进化
Initialize:
初始化 Pm, Pc, M, G, Tf 等参数;
随机产生第一代种群 Pop;
While  (个体得分未超过 Tf，或繁殖代数未超过 G)  do
计算种群 Pop 中每一个体的适应度 F(i);
初始化空种群 newPop;
Do
{
根据适应度以比例选择方法从种群 Pop 中选出两个个体;
if ( random(0,1) < Pc )
对两个个体按交叉概率 Pc 执行交叉操作;
if ( random (0,1) < Pm )
```

```
对两个个体按变异概率 Pm 执行变异操作；
将两个新个体加入种群 newPop 之中；
} until ( M 个子代被创建 )
用 newPop 取代 Pop；
End While
Output：最优解。
```

3. 算法特点

遗传方法具有自组织性、自学习性、自适应性和并行性；不需要求导计算或其他辅助知识，只要确定影响搜索方向的目标函数和对应的适应度函数；转换规则强调概率而非确定的；对给定问题可产生多个的潜存种，最终选择可由使用者确定，因此适合于多目标优化问题。

在遗传进化的初期通常会产生一些超常个体，如果采用比例选择法，异常个体因竞争力太突出而控制选择过程，将影响方法的全局优化进程。在遗传进化的后期，即方法接近收敛时，由于种群中个体适应度较小，继续优化的潜能降低，可能获得某个局部最优解。上述问题称为遗传方法的欺骗问题。当适应度函数设计不合理时有可能出现该问题。

当种群中个体适应度非常相似时，这些个体进入配对集的概率相当，而且交配后得到的新个体也不会产生较大变化。因此导致不能有效进行搜索，有可能趋向于纯粹的随机选择，从而使进化过程陷于停顿状态，无法找到全局最优解。针对该问题，可在迭代过程中用部分优质的新子个体来更新部分父个体作为下一代种群，即采用稳态繁殖。

第 5 章

大规模人脸搜索系统

在海量视频和图像数据中，可根据某个人的人脸图片、画像、监控人像、目击者描述等，快速查找出该人的相关视频和图像，然后获取到其姓名、单位、住址、微博、微信、爱好、亲友等关联信息，最后统计出他（她）的社会关系、日常行踪与活动轨迹，这就是大规模人脸搜索系统。

本章以“大海捞针”人脸搜索系统为例，首先介绍人脸检测、人脸特征提取、人脸特征比对等核心方法；然后详细阐述该搜索系统的体系结构、关键技术、实现流程、伪代码和性能评测方法等；最后介绍该系统的性能特点和使用方法。

5.1 概述

当去银行办理业务时，柜台服务员总是要求我们首先出示有效证件；当进入办公大门时，总是需要我们先刷卡或录指纹才能进门。在生活中，我们经常会遇到这种明明“我就在那里”却需要证明“我就是我”的事情，这个问题一直困扰着人们，无论是古老的签字画押，还是现代的身份证或通行证，都在解决一个问题：“我是谁”。

小明是个内向、单纯的小伙子，某天在咖啡馆里，小明对初次见面的女孩子很有好感，但是不知如何找到合适的话题。他很想知道她喜欢什么。在短暂的沉闷之后，他用手机拍下了女孩的照片，基于人脸搜索到她爱好旅游和偶像明星，小明马上找到了共同

话题，开始慢慢和对方侃侃而谈。

某市发生了一起刑事案件，犯罪份子逃之夭夭，“他在哪”困扰着人们。从现场监控录像中，办案人员得到嫌疑人的面部图像。经过在线监控系统的实时追踪，在海量的视频图像中，办案人员搜索到犯罪分子近期的活动地点，并很快将其绳之以法。

上述问题的解决都依赖于人脸搜索系统。如图 5.1 所示，无论在监控视频数据库中还是在社交网络上，都存在海量的人脸视频或图像。人脸搜索系统通过摄像机或视频监控设备等获取若干图像或视频片段，首先利用计算机对输入图像或视频进行人脸检测，搜索图像或视频中是否存在人脸并判断其位置和大小，提取出人脸面部图像；然后根据决策系统下达的任务指令进行识别，把人脸与身份信息对应起来，或者利用人脸对个体进行跟踪定位；接着与网络系统或者数据库相连接，搜索与该个体相对应的附属信息，如兴趣爱好等。决策系统通过干预整个搜索系统，对其进行反馈修正，指导输入图像或视频的选取与采集，如图 5.2 所示。人脸搜索系统涉及图像处理、模式识别、计算机视觉、统计学、人工智能、认知科学等多个领域，具有广泛的应用前景。



图 5.1　人脸数据库

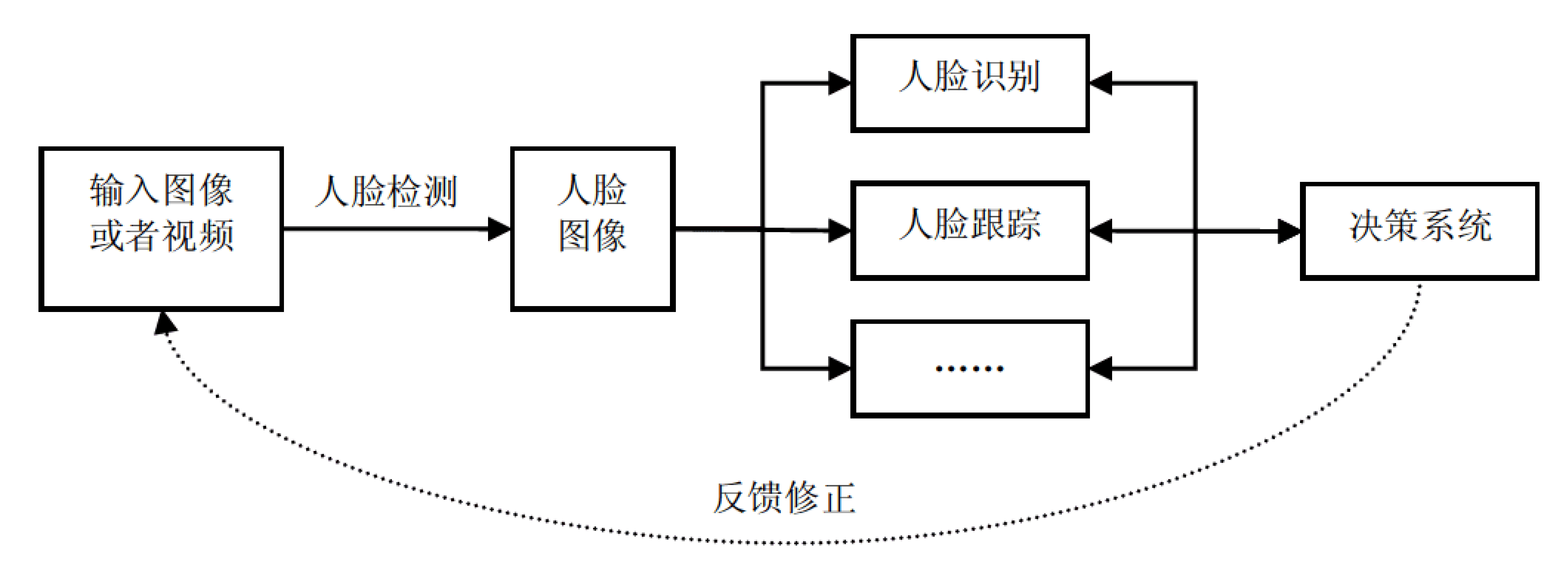


图 5.2　人脸搜索系统

为什么要以人脸为载体发展搜索系统呢？可以从下面几个方面来理解。俗话说“眼睛是心灵的窗户”，根据人眼的视觉特性，人们通常对图像中的人脸区域最感兴趣；人脸是人体的一个具有很强表征模式的内在属性，包含年龄、性别、表情等丰富的个体信息量，具有很强的自身稳定性和个体差异性，与个体高度结合、不会被遗忘或丢失；样本采集方便、设备成本低，较少或不需要个体的主动配合，易被用户接受，潜在的数据资源丰富。因此，基于人脸的搜索比传统的基于文字内容的搜索更实用，可以广泛地应用于身份认证、安全访问控制、视频监控、内容检索、人机交互、孤寡老人照料等各种领域，人们还可以通过人脸搜索查找特定的人脸图片、组织管理照片或者寻找与哪位名人最像，甚至可以利用上传的照片或视频在社交网络上发觉隐藏在图片或视频背后的社会关系等。

如图 5.3 所示，人脸搜索采用视频图像处理、模式识别、机器学习、视频分析等方法，通过图像预处理、镜头检测、关键帧提取、内容关联、视频语义化等技术，实现面向大数据的特定人脸目标查找与定位。

大规模人脸搜索技术改变了传统的视频信息组织方式，将海量视频数据按照时间、地点、来源等相关信息实现统一管理，便于查看、搜索与维护。将视频分析功能进行整合，用户使用更方便，系统维护与升级更容易。可根据用户需求，采用快速浏览、特征搜索等查看方式，不仅节约人力，而且自动处理方式能够有效克服人工搜索导致的疲劳漏查、效率低下等问题。

尽管需求性很强、应用范围很广、商业价值很高，然而目前没有成熟的人脸搜索系统。建立和发展人脸搜索系统具有如下困难。

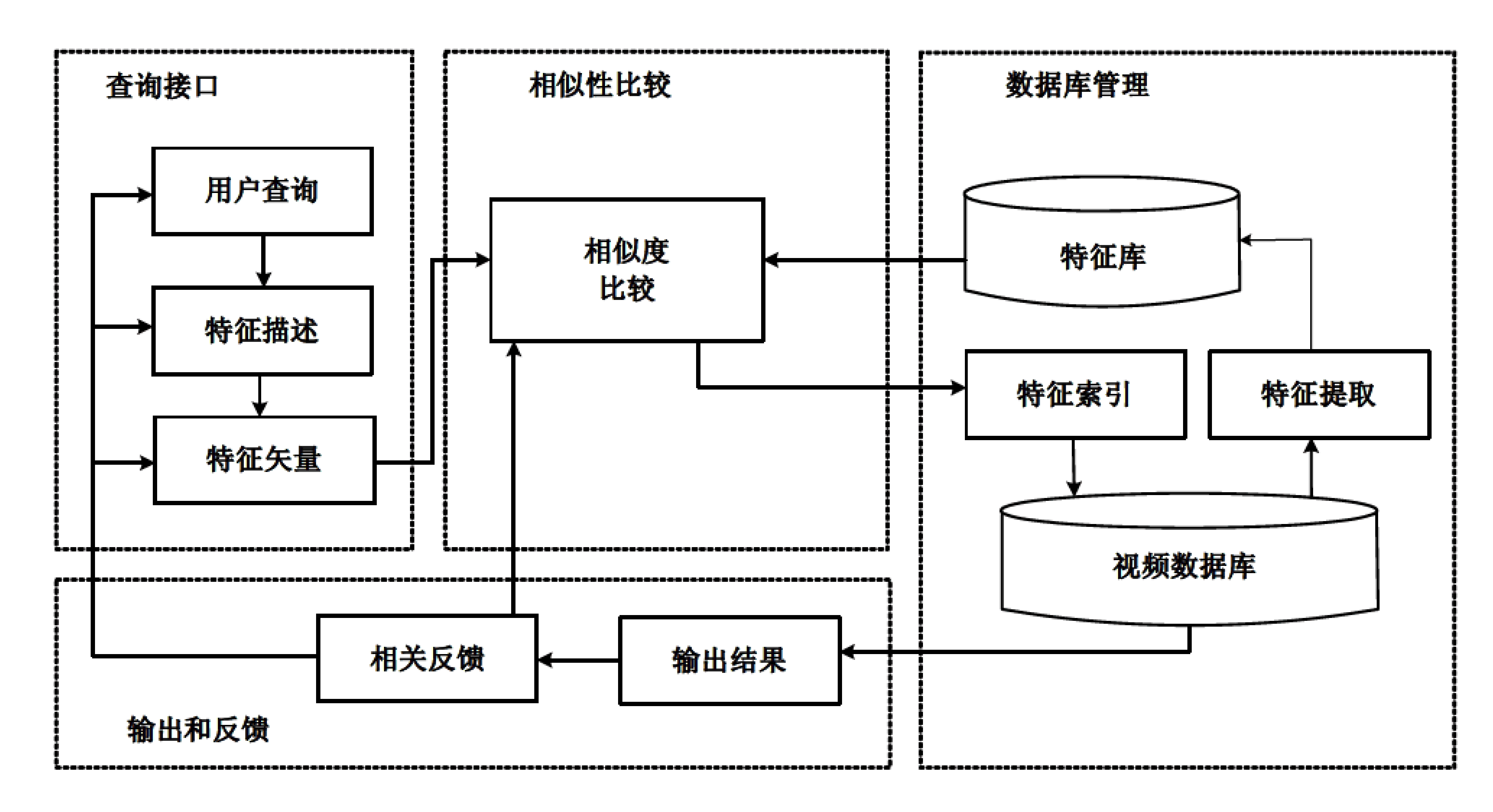


图 5.3　人脸搜索框架

1. 小样本问题

在人脸搜索的各个环节均存在严重的小样本问题。

在人脸检测环节，感兴趣的检测区域大多数是非人脸区域，人脸在这里是小样本，导致在训练过程中训练器逐渐向非人脸样本倾斜。

在人脸识别环节，某个个体的带身份标签信息的样本可能只有几个，而我们面对的待识别对象却可能是成千上万个，基于互联网的任何搜索面对的数据库都以亿为单位。

2. 内在复杂性

人脸在不同时刻会出现抬头、低头、转头、摇头、睁眼、闭眼等各种姿态，会由于心情变化而伴有高兴、发怒、痛苦、忧伤等各种表情，会随着年龄增长而有青春痘、皱纹，会因为整容使面部特征发生改变，这些都会导致人脸搜索系统的准确度下降。

3. 外在干扰

人脸本身的特征常常被胡须、眼镜、头发、帽子、围巾等附属物所遮挡，光照和环境条件的改变，摄像机或视频监控设备采集图像时电子原件噪声、人为抖动、姿态旋转等因素都可能影响图像质量，导致一人千面，给人脸搜索系统的泛化能力带来巨大挑战。

4. 搜索速度

任何一个系统如果不能在人类有限的耐心消耗完之前给出满意结果，都是不实用的。速度要求人脸搜索系统能对用户需求快速响应并返回结果，必须根据具体任务需求

权衡实时性和准确度之间的关系，在一定的精度下加快搜索速度。

国内外很多高校和研究机构都在从事人脸搜索系统相关的研究，如麻省理工学院（MIT）、卡内基梅隆大学（CMU）、南加州大学（USC）、清华大学、北京大学、亚洲微软研究院、中国科学院、Google、Facebook、百度等。相关论文在国际期刊和会议上的发表数量也逐年递增，如 IEEE 的 AFGR（Automatic Face and Gesture Recognition）、ICIP（International Conference on Image Processing）、CVPR（Computer Vision and Pattern Recognition）等会议上有近一半的论文都与人脸搜索有关，著名期刊 PAMI（Pattern Analysis and Machine Intelligence）在 1997 年 7 月和 2011 年 10 月出版了两期人脸识别专辑。

美国专利商标局于 2011 年 5 月发布了 Google 针对公众人物的人脸识别技术专利，即自动挖掘公众人物形象的视觉搜索应用。如图 5.4 所示，通过提供公众人物名单和其中某个待搜索的人员面部图像，该人脸搜索系统可以生成一个精确的与该面部图像一致的形象，然后按照一定的精确度判别该人脸属于哪个公众人物，或者是名单之外的人员。对人脸图片的搜索不仅返回一张相似照片，而且返回该目标人物在网络上的任何图片。该专利已经对 1000 幅图像进行过测试，测试图像包括美国总统奥巴马、流行歌手布兰妮、英国哈里王子、演员布拉德·皮特等。



图 5.4　Google 研制的人脸搜索系统

5.2 人脸检测

任何人脸搜索系统首先都需要从摄像机或视频监控设备采集得到的输入图像或视频中搜索出人脸及其位置和大小，人脸检测是系统的第一个关键所在，其获取人脸图像的精度与速度直接决定着整个系统的性能。

人脸检测（face detection）是指对于任意一幅给定的输入图像，采用一定的策略对图像进行搜索以确定其中是否含有人脸，如果有则返回人脸在图像中的数量及每张人脸的位置、大小和姿态。人脸检测是比人脸定位更宽泛、更复杂的技术，人脸定位（face location）一般是指在事先知道给定的输入图像中人脸数量（通常有且仅有一张人脸）的情况下去查找人脸所在的位置。人脸检测的示意图如图 5.5 所示。



（a）原始图像



（b）人脸检测图像

图 5.5 人脸检测示意图

5.2.1 人脸检测方法分类

如图 5.6 所示，从方法论上讲，人脸检测方法可以根据不同的准则进行不同的分类，基本上可以归纳为以下几种。

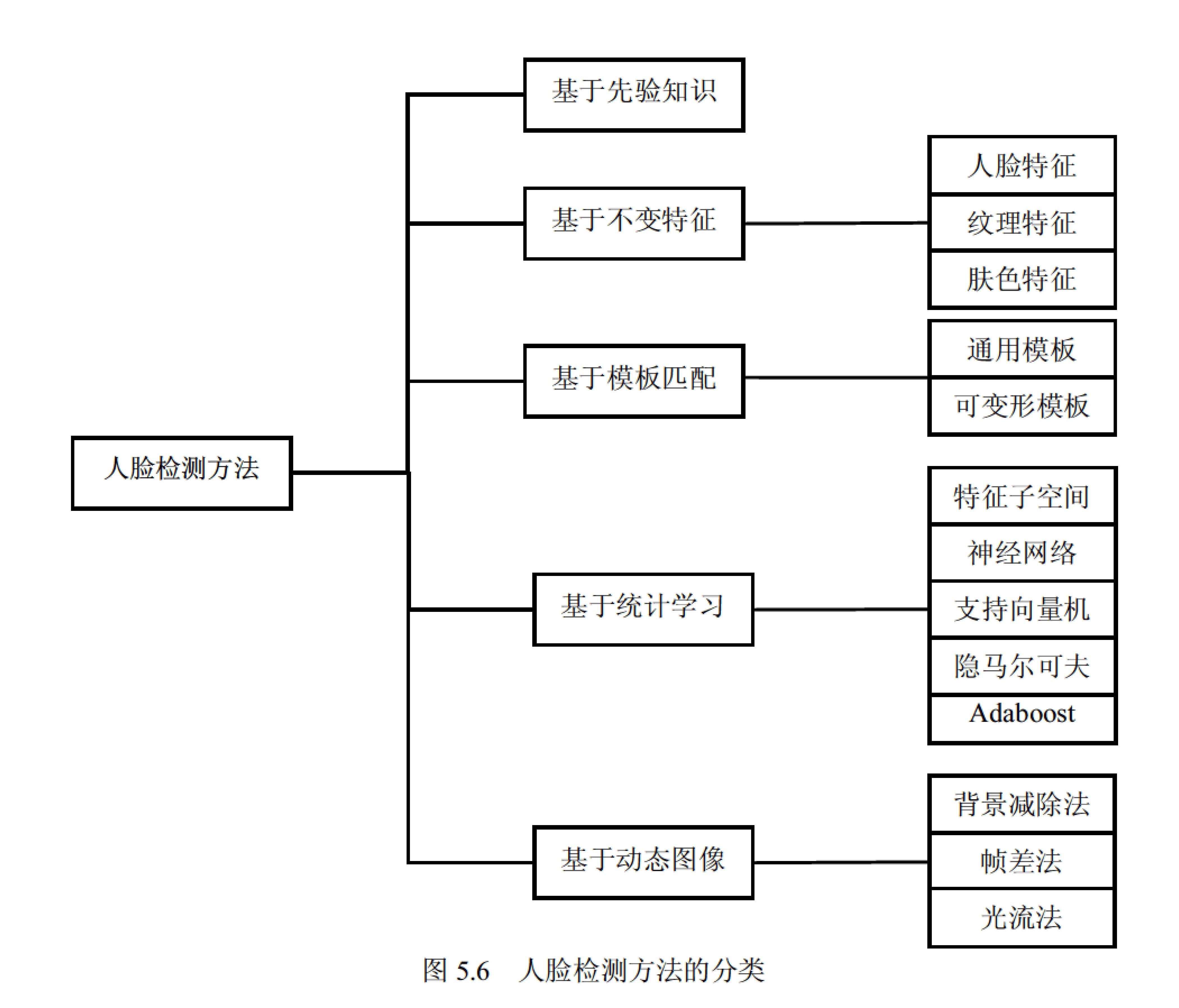


图 5.6　人脸检测方法的分类

1. 基于先验知识的方法

这类方法主要基于人们对典型人脸面部特征之间相互关系的认识制定出一系列的规则来检测人脸，如鼻子位于嘴巴上面、两只眼睛的连线与面部中轴线垂直等。但是，精确恰当的判定规则的定义是非常困难的，规则太细则检测时很难满足所有规则；反之，规则太粗则可能得到太多的伪人脸。因此，这类方法主要用于正面人脸图像和人脸定位，属于早期的自上而下的人脸检测方法。

2. 基于不变特征的方法

这类方法的目标是找出在姿态、视角、光照变化的情况下仍然保持不变的人脸结构特征来定位人脸。这些不变的结构特征包括人脸特征（如眼睛、眉毛、嘴巴、鼻子）、纹理特征、肤色特征等以及它们的组合特征。这类方法是自下而上的方法，对姿态等变化不敏感、检测过程相对稳定。

程序复杂度低，但当背景图像比较复杂时检测精度会下降，主要用于人脸定位。

3. 基于模板匹配的方法

这类方法使用预先存储的一些描述整张人脸或人脸局部特性的标准人脸模板，通过计算待检测图像与存储模板之间的相似性大小来进行匹配检测。在实际应用中，所采用的人脸模板可以分为通用模板和可变形模板。这类方法既可以用于人脸器官的精确定位，也可以用于人脸配准，其优点是简单直观、容易实现，缺点是很难有效地处理搜索尺度、姿态和光照变化等问题。

4. 基于统计学习的方法

这类方法将人脸区域和非人脸区域看成两种不同的模式，从而将人脸检测问题转化为模式识别中的“两类”分类问题，通过利用某种统计分析或机器学习方法对大量的人脸样本与非人脸样本进行训练以得到它们各自的统计特征，继而解析出一个人脸模型并构建分类器，然后使用训练得到的分类器判断输入图像中所有感兴趣区域属于哪类模式，以此完成人脸检测。

具有代表性的方法包括特征子空间方法、神经网络方法、支持向量机方法、隐马尔可夫方法、Adaboost 方法等。这类方法具有很强的适应能力、实时性和鲁棒性，是目前最流行，也是成就和影响最大的方法，适用于复杂背景图像中的实时人脸检测，其缺点是需要大量的训练样本和统计分析，训练过程费时费力，故仍有待改进。

5. 基于动态图像的方法

随着时代的发展，视频监控设备随处可见，这些设备能得到一段时间内连续的若干幅动态图像，即视频序列。这些连续的视频图像中包含了更多的时间相关信息、运动信息和前后帧关联性。这些额外的信息促发人们探索有别于上述基于静态图像的人脸检测方法，将人脸从复杂的背景中有效地分割出来，比较成熟且应用广泛的方法有背景减除法、帧差法、光流法等。

5.2.2　基于 Adaboost 的人脸检测

在实际应用中，人脸检测常使用 Adaboost 方法，Adaboost 方法属于基于统计学的方法，能在达到较高检测精度的同时快速返回检测结果。

1. 基本原理

Viola 和 Jones 提出的基于 Adaboost 的人脸检测方法有机地组合了 3 个重要思想：

- 使用 Haar-like 特征表示图像，引入积分图，提高特征计算速度；
- 基于弱分类器，采用 Adaboost 方法，选择少量特征构造强分类器；
- 使用级联（Cascade）策略提高人脸检测速度。

相关论文为 2001 年发表在 CVPR 上的 *Rapid object detection using a boosted cascade of simple features*，被引用的次数已经超过 8000 次。由于在人脸检测领域的基础性影响和里程碑意义，该论文获得 2011 年 CVPR 委员会颁发的 Longuet-Higgin 奖，表彰该文章 10 多年来在计算机视觉领域做出的奠基性贡献。

2. Haar-like 特征与积分图

Haar-like 特征和积分图（integral image）是 1984 年由富兰克林·克罗引入计算机图形学领域的，但是该概念并没有在计算机图形学领域被广泛应用，近 20 年后因在 Viola 和 Jones 的人脸检测框架中取得成功开始广受关注。

Haar-like 特征是一种基于 Haar 小波的特征，最早由 Papageorgiou 等应用于人脸表示。采用 Haar-like 特征代替常用的图像强度特征（即图像中每个像素点的 RGB 值）的原因在于后者的计算量很大，而前者与积分图的结合可以实现对特征的快速计算，任意尺寸的 Haar-like 特征都可以在常数时间内完成。

Haar-like 特征使用检测窗口中指定位置的相邻矩形，计算每个矩形区域的像素和，并取其差值来对图像的子区域进行分类。如图 5.7 所示，在 Viola 和 Jones 提出的人脸检测框架中，Haar-like 特征可分为 3 类和 4 种形式，即 2 矩形特征（2-rectangle features，子图 A 和 B）、3 矩形特征（3-rectangle features，子图 C）和 4 矩形特征（4-rectangle features，子图 D）。更多的 Haar-like 特征（如倾斜 45°的特征）可参考 Lienhart 和 Maydt 的文献。每种形式的 Haar-like 特征可以具有任意的位置和尺寸，但均包含白色和黑色两种矩形，其特征值为白色矩形的像素和减去黑色矩形的像素和。Haar 特征值反映图像中特定区域的某些特性，比如边缘或者纹理，其值随特征形式、矩形大小和矩形位置的变化而变化。

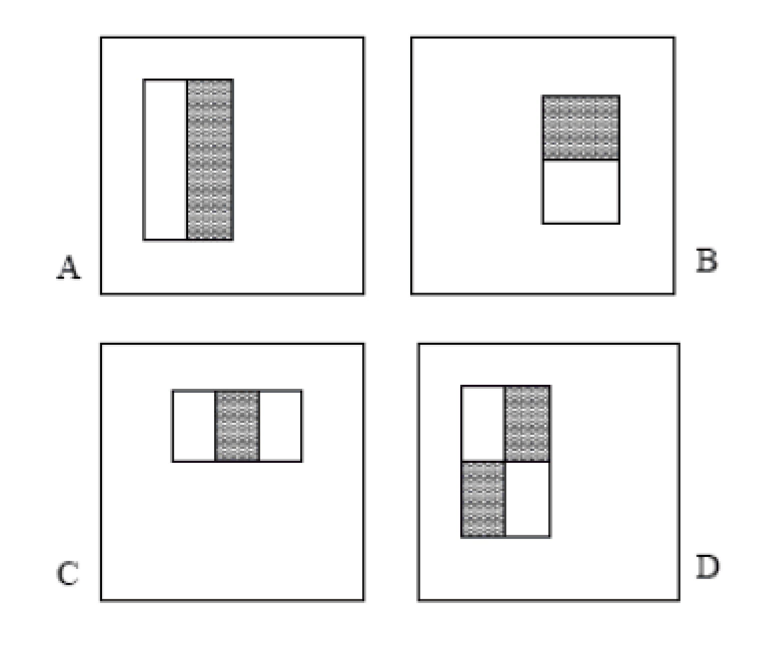


图 5.7 部分 Haar-like 特征

在一个很小的检测窗口中，可能包含非常多的矩形特征，如在 24×24 像素大小的检测窗口中，矩形特征的数量可以达到 18 万个，远远大于检测窗口中的像素总数 24×24=576。如何快速计算这么多的特征就成为一个迫切的问题，积分图可以解决该难题。

积分图是描述图像全局信息的一种矩阵表示方式，其每一点的值是原始图像中对应位置的左上角区域的所有值之和：

$$I(x,y)=\sum_{x'\le x,y'\le y} i(x',y')$$

其中，$I(x,y)$ 是积分图像，$i(x',y')$ 是原始图像的像素值。

整个图像的“积分图”只需要遍历一次图像就可以全部计算出来，这可以从下面的关系式得出：

$$I(x,y)=i(x,y)+I(x-1,y)+I(x,y-1)-I(x-1,y-1)$$

推广上述关系式，可以在常数时间内计算出图像中任意矩形区域的像素值之和。如图 5.8 所示，阴影矩形区域的值为：

$$\sum_{\substack{x_1\le x\le x_4\\ y_1\le y\le y_4}} i(x',y')=I(4)+I(1)-I(2)-I(3)$$

即矩形特征的特征值计算只与该矩形端点的积分图有关，而与图像坐标值无关，并且整个计算过程只需要进行简单的加减运算，不管矩形特征的尺度如何，其特征值所需的计算资源是常量。因此积分图的引入使得 Haar-like 特征的计算更加方便、快速，为实时人脸检测提供了保证。

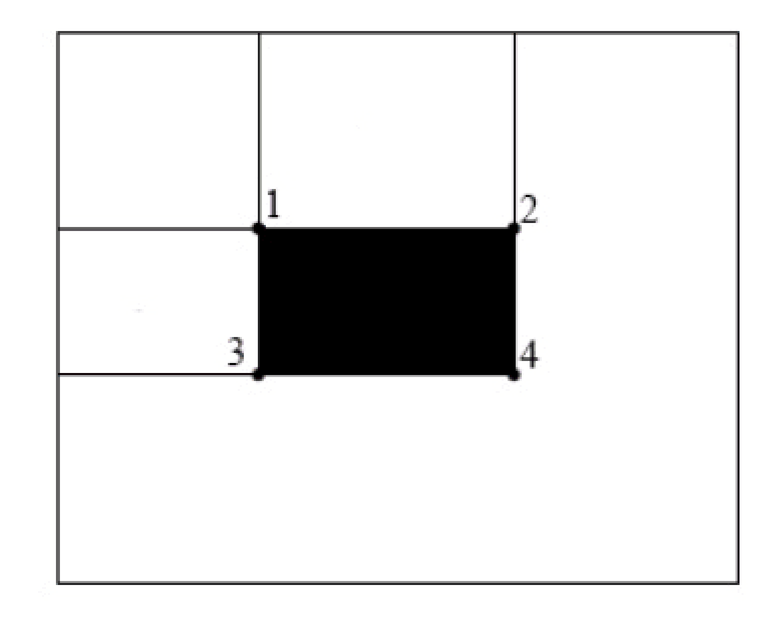


图 5.8　积分图计算矩形阴影区域的值

3. Adaboost 方法

在积分图解决快速计算大量矩形特征的特征值问题之后，需要采用 Adaboost 方法来训练分类器，使得 Haar-like 特征恰到好处地组合起来以检测人脸。

Adaboost（Adaptive Boosting，自适应增强）是一种统计学习方法，由 Freund 和 Schapire 于 1995 年提出，其基本思想是将大量分类能力一般的弱分类器，通过一定方式组合起来，如“绝对多数”投票或加权投票，构造一个分类能力很强的强分类器。所谓弱分类器（weak classifier）并不局限于某种确定类型的分类器，“弱”只是表明分类器的分类能力不是很强，即只需要分类器的精确度比随机猜测稍微有些提升即可，如对于两类问题而言，分类正确率只需超过 50%，而强分类器的精确度要求远远超出随机猜测的精度。在 Viola 和 Jones 的人脸检测框架中，一个 Haar-like 特征对应着一个弱分类器 $h_j(x)$，其定义为：

$$h_j(x)=\begin{cases}1, & p_j f_j(x)<p_j\theta_j \\ 0, & \text{其他}\end{cases}$$

其中，$h_j(x)$表示弱分类器的值，1 表示人脸，0 表示非人脸；x 表示一个待检测的子窗口，p_j 用于控制不等式的方向，即只能取 ±1；$f_j(x)$为某个 Haar-like 特征的特征值，θ_j 为阈值。

在生成弱分类器之后，如何从弱分类器组合中得到强分类器呢？Adaboost 方法的基本思想是：通过不断迭代，自适应地调整弱分类器的错误率，直到错误率能达到某个预定的足够小的期望值。

Adaboost 方法首先对每个训练样本赋予一个权值（初始权值为常数），在每一轮迭代时，根据分类结果对前一轮训练失败的样本赋以较大的权值，而对正确分类的样本则降低其权值；然后让学习算法在后续学习中“聚焦于”这些比较难分的训练样本之上；最后由算法挑选出若干个弱分类器，加权相加组成强分类器。

4. 级联分类器

通过 Adaboost 方法可以从弱分类器中训练合成强分类器，提升分类器的精确度。在现实的人脸检测中，仅仅靠一个强分类器难以保证检测的正确率。需要采用级联（Cascade）策略将训练出的多个强分类器“强强联手”形成级联分类器，在比较好地排除非人脸样本的情况下，提高人脸检测的正确率。

级联分类器的基本思想是：通过各级强分类器检测的对象为真实人脸的可能性会比较大。

如图 5.9 所示，级联策略将若干个强分类器分级串联在一起，在检测过程中，通过前几级分类器拒绝大量的非人脸样本（如背景区域），以节约更多时间专注于对那些更可能是人脸的区域进行检测，使得整个人脸检测的速度大幅提高。串联的强分类器需要一级比一级复杂，一级比一级包含的弱分类器多，以确保在最快的时间内排除最多的非

人脸样本，并逐级提高检测精度。

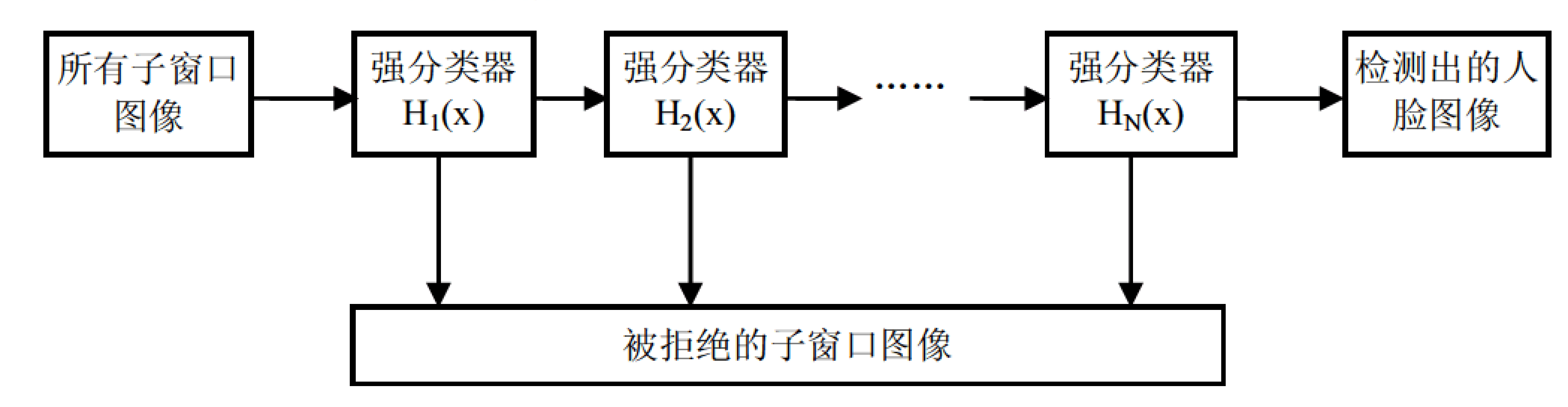


图 5.9 级联分类器的检测示意图

5.3 人脸特征提取

从海量的图像或视频中检测出人脸，仅仅完成了第一步；检测出的原始人脸图像通常包含的特征数量有成千上万个，容易引起“维数灾难”，高维特征使得训练模型更复杂、泛化能力下降；使得分析特征、训练搜索系统所需时间大大增加，难以满足人脸搜索系统的快速要求。

要获取有用信息，必须对人脸图像进行特征提取，以便减少特征个数、降低运行时间、得到标识人脸个体的最具有代表性的特征或最有利于模式分类的特征，如图 5.10 所示。

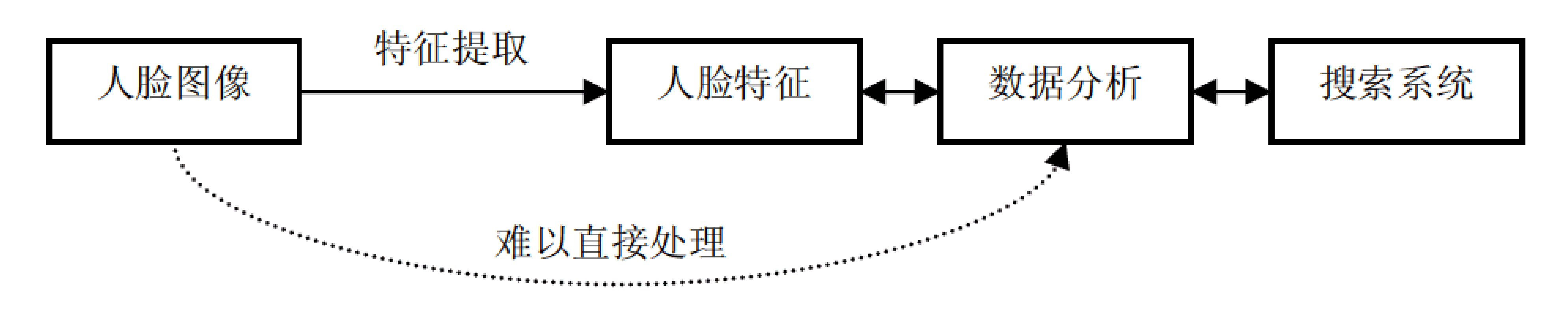


图 5.10 人脸特征提取

人脸特征提取就是用尽可能少的特征（如耳朵、眼睛、嘴巴等的位置、关键点或轮廓线）来紧凑地表示人脸，减少或消除人脸图像中的次重要信息，保持或突出足够满足人们需求的有效信息，便于后续的分析处理，如识别、跟踪、搜索关联信息等。在人脸识别中，一幅 64×64 的人脸图像按行或列堆叠最后转化为 4096 维像素空间中的向量，已有研究表明，可用降维后的子空间来有效表示不同光照条件下人脸图像模型。

人脸特征的提取方法可以根据不同准则进行分类。

1. 根据特征的组合方式

- 特征选择方法（feature selection）

人脸特征提取得到的特征是原始特征的一个子集，不改变原始特征的值。首先从所有人脸特征中，启发式或随机地搜索出一个特征子集；然后采用某个评价函数对其进行评价，以判定该特征子集的好坏程度，若评价结果比设定的准则好，就完成特征提取过程，否则继续搜索下一组特征子集。

❑ 特征抽取方法（feature extraction）

人脸特征提取得到的特征是原始特征的一个组合或一个映射，通常不再保持原始特征的值。首先根据某个假设建立相应的数学模型，并得到其目标函数；然后通过优化该目标函数得到所期望的特征。

2. 根据数据的结构特征

❑ 子空间方法

该类方法假设数据位于或近似位于低维的线性或仿射子空间上，采用具有显示表达式的线性映射函数提取数据特征，如主成分分析（Principal Component Analysis, PCA）。

❑ 流形学习方法

该类方法假设数据位于或近似位于某个非线性流形上，或者说数据具有全局的非线性分布和局部的近似线性分布，通过隐含映射方式提取出原始高维人脸特征的低维表示。其映射函数是非线性映射，通常没有显式表达形式，如局部线性嵌入（Locally Linear Embedding，LLE）。

3. 根据先验标签信息的多少

❑ 无监督方法

当无法获取数据的先验标签信息时，无监督方法试图在提取后的特征空间中尽量忠实地保持数据的全局或局部几何结构，以挖掘数据中隐含的有意义的特征，或对数据做一些探索性分析，如聚类等。

❑ 半监督方法

当大量的具有标签信息的数据难以获取时，半监督方法试图利用少量的标签数据和大量的无标签数据来提取特征，根据有标签数据的结果来设计分类器，从而更好地对无标签数据进行分类。

❑ 有监督方法

有监督方法假设训练样本的标签信息已知，充分利用标签信息来对训练样本进行学习，以选取出最有利于对训练样本集外的数据进行标记的人脸特征，如线性判别分析（Linear Discriminant Analysis, LDA）。

5.3.1 PCA 方法

PCA（Principal Component Analysis，主成分分析）是一种无监督的子空间特征抽取方法，由 Hotelling 于 1933 年首先提出，是最古老、最经典的数据分析工具之一。Turk 和 Pentland 于 1991 年将 PCA 方法应用于人脸表示和人脸分类（即特征脸，Eigenfaces），取得了成功，相关论文被引用次数已达 11000 余次。

在数据分析中，通常会假设数据分布服从一定的概率，如正态分布等。概率分布有两个关键的评价指标，即均值和方差。方差度量随机变量与其均值之间的偏离程度。某一维度上的方差衡量数据在该维度上的波动情况，方差越大，与均值的离散程度越大，所提供的有价值信息越丰富。

通常方差越大提供的信息越多，方差越小提供的信息越少。主成分分析以方差大小来作为信息量多少的依据，通过线性变换降低数据维数，其基本思想是尽可能地保留较大方差的数据，丢掉方差较小的数据，从而在尽量保留原始数据主要信息、损失最少有用数据的前提下提取特征，抓住事物的主要矛盾，并揭示数据内部的规律性。

给定一组观测数据 $\{x_i \in R^D, \mathrm{i}=1,2,\cdots,\mathrm{N}\}$，主成分分析的目标可以表述为：寻找一组相互正交的投影，或一个列正交的线性投影矩阵 $G \in R^{D\times d}$，使得投影后的低维数据表示 $y_i = G^T x_i$ 具有最大的方差。

记原始数据按列堆叠构成的矩阵为：

$$X = \left[x_1, x_2, \cdots, x_N\right] \in R^{D\times N}$$

低维嵌入表示按列堆叠构成的矩阵为：

$$Y = \left[y_1, y_2, \cdots, y_N\right] \in R^{d\times N}$$

原始数据的样本协方差矩阵为：

$$S_t = \sum_{i=1}^{N}(x_i - \overline{x})(x_i - \overline{x})^T = XHX^T \text{ [1]}$$

其中，$\overline{x} = \sum_{i=1}^{N} x_i / N$ 为样本均值，$H = I - ee^T / N$ 为中心化矩阵，I 是单位矩阵，$e \in R^N$ 是元素全为 1 的列向量。

由 $Y = \left[y_1, y_2, \cdots, y_N\right] = [G^T x_1, G^T x_2, \cdots, G^T x_N] = G^T X$ 可求得低维嵌入表示的协方差矩阵为：

[1] 注意，严格的样本协方差矩阵定义应为 $S_t / (N-1)$，这里为方便省去了乘积因子 $1/(N-1)$，但这不会影响后续的分析和结果，下同。

$$\sum_{i=1}^{N}(y_i-\overline{y})(y_i-\overline{y})^T=YHY^T=G^TXHX^TG=G^TS_tG$$

其中，$\overline{y}=\sum_{i=1}^{N}y_i/N$ 为低维嵌入表示的均值。

主成分分析的目标函数可以表示为下列数学形式：

$$\begin{aligned}&\arg\max_{G}\quad tr(G^TS_tG),\\&s.t.\qquad G^TG=I,\end{aligned}$$

其中，$tr(G^TS_tG)$ 表示矩阵 G^TS_tG 的迹，s.t.是 Subject to 的缩写。

上述目标函数的最优解，即主成分分析对应的变换矩阵 G 可以通过对协方差矩阵 S_t 进行谱分解或特征分解来求解。

定理 5.1 主成分分析的最优线性变换矩阵 G 由 S_t 的最大 d 个特征向量组成。

证明：使用拉格朗日乘子法来最大化上述目标函数，得该优化问题的拉格朗日函数为：

$$L(G,\lambda)=tr(G^TS_tG)-tr(A(G^TG-I))$$

其中，A 为实对称矩阵。

对 G 求偏导，令结果为 0，可得

$$\frac{\partial L(G,\lambda)}{\partial G}=S_tG-GA=0$$

由线性代数理论，实对称矩阵可用正交矩阵对角化。因此，存在正交矩阵 U 和对角矩阵 $\Lambda=diag(\lambda_1,\lambda_2,\cdots,\lambda_d)$，使得 $A=U\Lambda U^T$。可得：

$$S_tG=GU\Lambda U^T\Rightarrow S_tGU=GU\Lambda$$

令 $GU=\Phi=\left[\phi_1,\phi_2,\cdots,\phi_d\right]$，则由上式有：

$$S_t\phi_i=\lambda_i\phi_i,i=1,2,\cdots,d$$

说明 λ_i 和 ϕ_i 是 S_t 的特征值和特征向量。

由目标函数及其约束条件，可知

$$\begin{aligned}tr(G^TS_tG)&=tr(G^TGA)=tr(A)=tr(U\Lambda U^T)\\&=tr(U^TU\Lambda)=tr(\Lambda)=\lambda_1+\lambda_2+\cdots+\lambda_d\end{aligned}$$

即目标函数值实际对应 S_t 的 d 个特征值之和，为使该目标函数值最大化，应取 S_t 的最大 d 个特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$。这样，$\Phi = [\phi_1, \phi_2, \cdots, \phi_d]$ 由 S_t 的最大 d 个特征值对应的特征向量组成。进而，有 $G = \Phi U^T$，即主成分分析的最优解 G 由样本协方差矩阵 S_t 的最大 d 个特征值对应的特征向量右乘一个正交矩阵 U^T 组成。

由于 G 右乘一个任意的正交矩阵不影响目标函数中优化问题的解，因此通常取最优线性变换矩阵 G 为 S_t 的最大 d 个特征向量。

图 5.11 给出主成分分析的一个例子，可以看出主成分分析对具有本征线性结构的数据集具有较好的分类效果，能通过学习获得均方误差下的最佳线性投影方向。

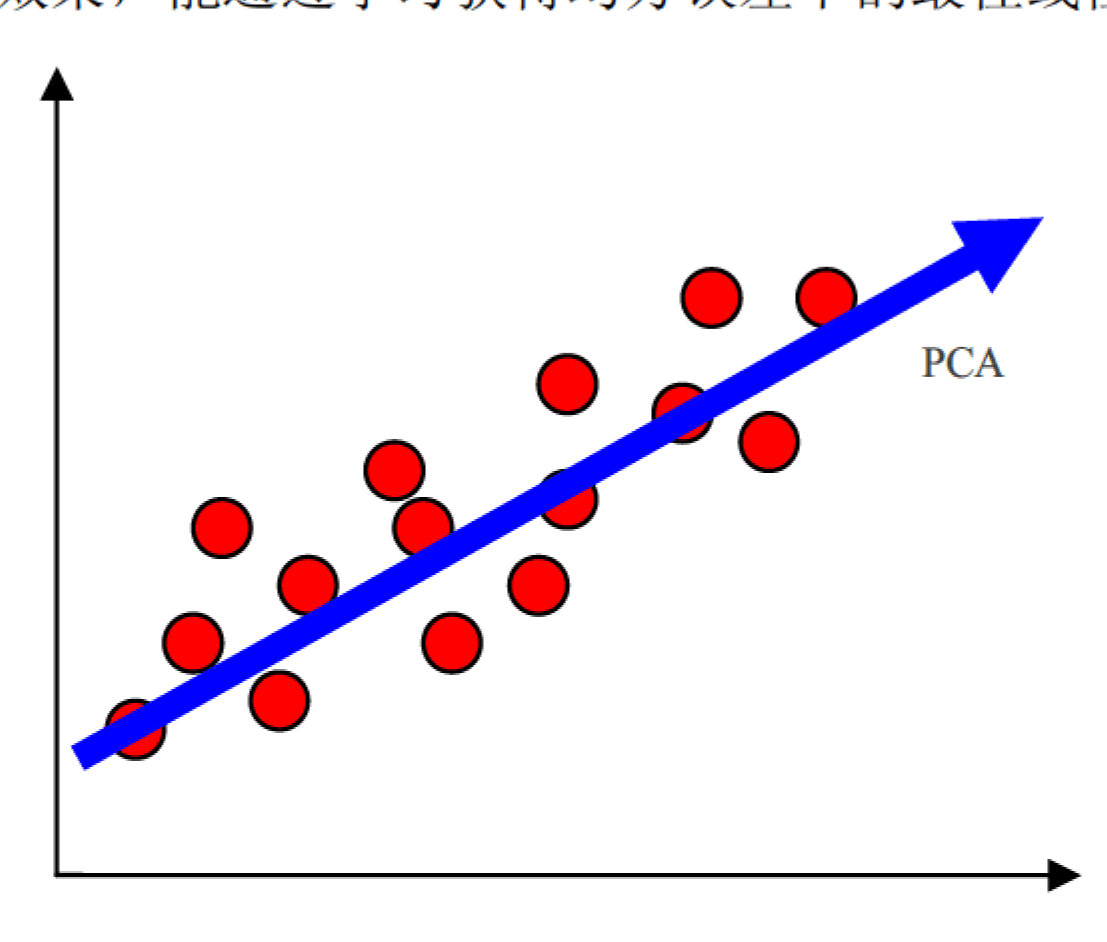


图 5.11　主成分分析方法示意图

5.3.2　LDA 方法

LDA（Linear Discriminant Analysis，线性判别分析）最早由 Fisher 于 1936 年提出，最初是为了解决两类数据的分类问题，后来被推广到多类数据的分类，并逐渐得到广泛关注和应用，成为特征提取算法中最经典的有监督方法之一。Belhumeur 等在 1997 年将 LDA 方法应用于人脸识别之中（即 Fisher 脸，Fisherfaces），从此 LDA 成为人脸识别的一种基准（baseline）方法，被引用次数已近 8000 次。

线性判别分析直接以数据分类为目标，基于这样一个直觉启发，即如果同一类的样本之间相对聚集，而不同类的样本之间相对远离，则不同的类能尽可能地分离，此时也能更容易分类识别出不同类的样本。

数据方差刻画样本波动的大小及其离散趋势，因此线性判别分析方法试图寻求一个低维的特征空间，使得类内离差尽可能小，而类间离差尽可能大。

给定来自 k 个类别的共 N 个数据采样 $x_1^1,\cdots,x_{N_1}^1,\cdots,x_1^k,\cdots,x_{N_k}^k \in R^D$，其中 x_j^i 是来自第 i 个类别的第 j 个样本，则有：

$$N = \sum_{i=1}^{k} N_i$$

不失一般性，假设这些数据按类别顺序封装在一个 $D \times N$ 的数据矩阵 $X = \left[x_1^1,\cdots,x_{N_1}^1,x_1^2,\cdots,x_{N_2}^2,\cdots,x_1^k,\cdots,x_{N_k}^k\right]$ 中，即最前面是来自第一类的 N_1 个样本，接着是来自第二类的 N_2 个样本，直到最后一类样本。线性判别分析的目标可以表述为寻找一个线性投影矩阵 $G \in R^{D\times d}$，使得经过特征提取后，属于同类的数据尽量靠近，而属于不同类的数据之间尽量远离。

原始数据的类内散布矩阵（within-class scatter matrix）S_w 和类间散布矩阵（between-class scatter matrix）S_b 分别为：

$$S_w = \sum_{i=1}^{k}\sum_{j=1}^{N_i}(x_j^i - \overline{x}^i)(x_j^i - \overline{x}^i)^T$$

$$S_b = \sum_{i=1}^{k} N_i(\overline{x}^i - \overline{x})(\overline{x}^i - \overline{x})^T$$

其中，$\overline{x}^i = \sum_{j=1}^{N_i} x_j^i / N_i$ 为第 i 类样本的均值，而 $\overline{x} = \sum_{i=1}^{N} x_i / N$ 为所有样本的均值。

线性判别分析的目标函数可以表示为下列数学形式：

$$J(G) = \underset{G\in R^{D\times d}}{\arg\max}\quad tr\left((G^T S_w G)^{-1}(G^T S_b G)\right)$$

其中，$G^T S_w G$ 和 $G^T S_b G$ 分别为特征空间中的类内散布矩阵与类间散布矩阵。

上述目标函数的最优解，即线性判别分析对应的变换矩阵 G 可以通过对原始数据的类内散布矩阵 S_w 与类间散布矩阵 S_b S_b 形成的广义特征值问题进行特征分解来求解。

定理 5.2　线性判别分析的最优线性变换矩阵 G 由广义特征值问题 $S_b g = \lambda S_w g$ 的最大 d 个特征向量组成。

当类内散布矩阵 S_w 非奇异时，可以通过对 $S_w^{-1} S_b$ 进行特征分解来得到线性判别分析的 d 个特征向量。在人脸搜索的应用中，类内散布矩阵 S_w 通常是奇异的，因此需要特别 fb 处理来避免对奇异矩阵的求解。

定理 5.3　在线性判别分析中，类内散布矩阵 S_w 的秩小于 $\min(D, N-k)$，而类间散布矩阵 S_b 的秩小于 $\min(D, k-1)$。

由于 $S_w \in R^{D\times D}$，$S_b \in R^{D\times D}$，由定理 5.3 可知，当数据的维数 D 大于样本数 N 时（人脸搜索中经常面对“小样本”问题，通常 $N \ll D$），类内散布矩阵 S_w 为奇异矩阵。为避免对奇异矩阵的求解，需要采用其他手段提取原始数据的特征，使得新的特征空间中的类内散布矩阵非奇异。在人脸搜索领域，通常采用主成分分析方法对原始特征进行预提取，得到非奇异的类内散布矩阵，然后通过在中间维数空间求解广义特征值问题，得到线性判别分析的投影矩阵。在人脸识别中，Fisher 脸(Fisherfaces)方法就是采用这种处理策略。

相对于主成分分析而言，线性判别分析最大的优点在于，直接以分类为目标，得到的投影方向是最能判别数据类别的方向，而主成分分析得到的是最能表达或重构数据的投影方向，如图 5.12 所示。

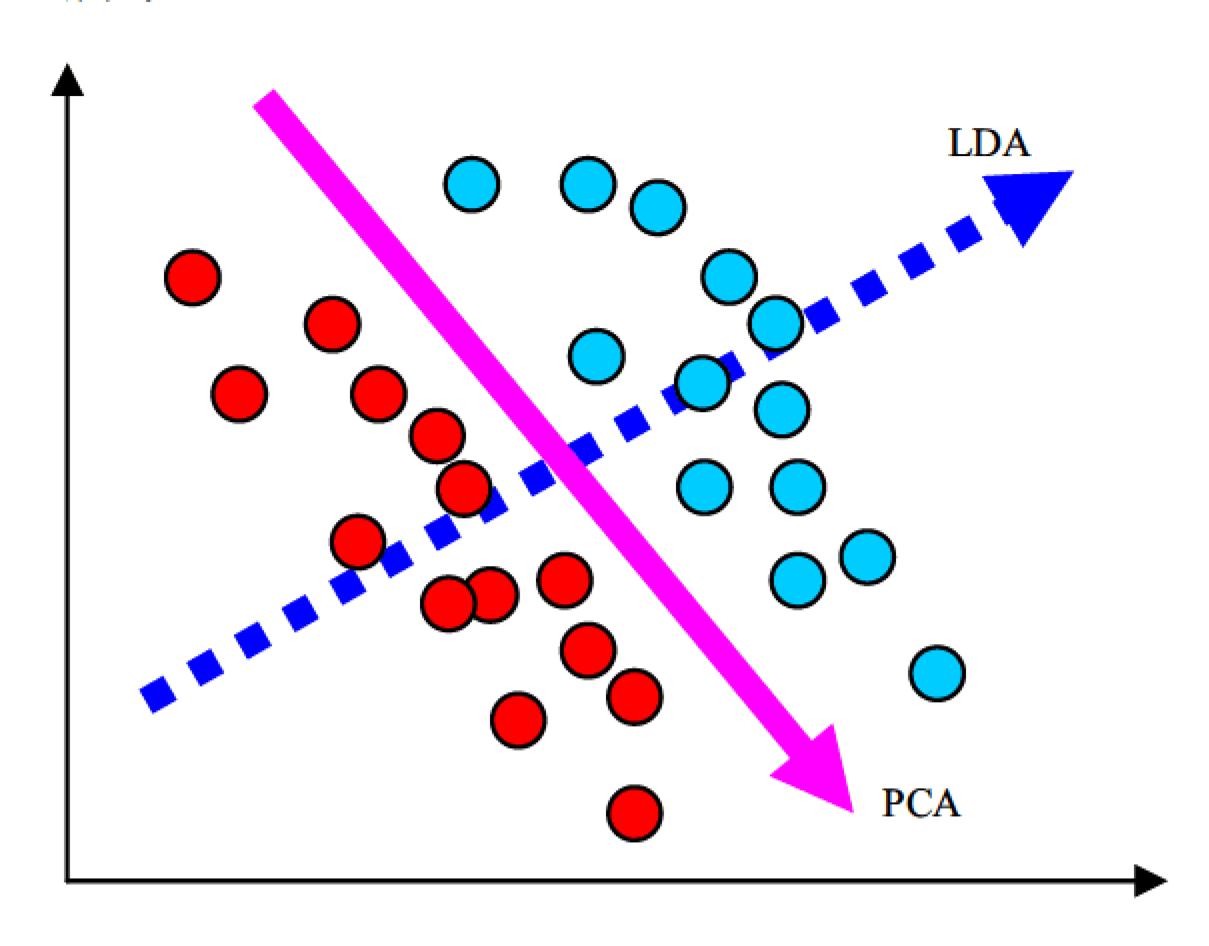


图 5.12　线性判别分析方法示意图

5.3.3　Kernel 方法

在数据位于低维线性或仿射子空间中时，传统的子空间特征提取方法处理效果很好，但是当数据位于非线性结构上时，会造成很大的扭曲，影响处理效果。

传统子空间特征提取方法大多为数据间的点积关系，为它们在核（Kernel）框架下进行扩展成为可能。

核化扩展的基本思想是：首先将数据投影到更高维的核空间，使数据近似地满足线性要求；然后在核空间直接采用经典的子空间方法有效地提取特征，如核主成分分析（Kernel PCA，KPCA）、核线性判别分析（Kernel LDA，KLDA）等。子空间方法的核化扩展已经成功应用于人脸检测、人脸识别和语音识别之中。

不妨假设数据已经中心化，即 $\sum_{i=1}^{N} x_i = 0$。可得：

$$Cu = \lambda u$$
$$C = \sum_{i=1}^{N} x_i x_i^T$$
$$\lambda \geq 0,\ u \in R^D \setminus \{0\}$$

由于

$$Cu = \sum_{i=1}^{N} (x_i^T u) x_i$$

因此所有对应 $\lambda \neq 0$ 的特征向量 u 必然位于 $x_1, x_2, \cdots, x_N$ 所组成的空间中，即存在系数 $\alpha_i, i = 1, 2, \cdots, N$ 使得

$$u = \sum_{i=1}^{N} \alpha_i x_i$$

另一方面，有：

$$x_k^T C u = \lambda (x_k^T u), k = 1, 2, \cdots, N$$

可得

$$\sum_{i=1}^{N} \alpha_i (x_k^T \sum_{j=1}^{N} x_j)(x_j^T x_i) = \lambda \sum_{i=1}^{N} \alpha_i (x_k^T x_i), k = 1, 2, \cdots, N$$

定义一个 $N \times N$ 的矩阵 K，其中元素为：

$$k_{ij} = k(x_i, x_j) \triangleq (x_i^T x_j)$$

则可以简化为

$$K^2 \alpha = \lambda K \alpha$$

其中，$\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_N]^T$。上式的求解归结为寻找下述特征问题的非零特征值和特征向量：

$$K \alpha = \lambda \alpha$$

求得 α 后，即可得到 PCA 的特征向量 u。

由于 PCA 的目标函数可以等价地重塑为数据间的点积关系，因此可以在核技巧的

框架下对其进行非线性扩展，得到核 PCA。核 PCA 将数据投影到一个更高维（可能是无限维）的点积空间Θ，该空间隐含地和一个非线性映射 ϕ 相关联：

$$\phi: R^N \to \Theta, x \mapsto \Phi(x)$$

使得数据近似地在更高维空间满足线性要求，此时可利用线性子空间学习方法 PCA 来处理数据。核 PCA 可归结为类似于下述特征问题：

$$K^* \alpha = \lambda \alpha$$

其中，K^* 是一个核矩阵，其元素为：

$$k_{ij}^* = \left(\phi(x_i)^T \phi(x_j)\right)$$

引入核矩阵而不是直接寻找非线性映射，计算更容易，不会比线性条件下增加多少额外的计算量。

常用的核矩阵如下。

- 高斯核

$$k^*(x_i, x_j) = \exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma^2\right)$$

- d 阶多项式核

$$k^*(x_i, x_j) = (x_i^T x_j + R)^d$$

- 线性核

$$k^*(x_i, x_j) = x_i^T x_j$$

其中，高斯核可以将原始空间映射为无穷维空间，对于一个特定的问题，不同的核函数可能会带来不同结果，在实际应用中需要通过尝试来得到或者需要一些经验信息。

5.4　人脸特征比对

将提取到的人脸特征与数据库中已有人脸特征进行比对，找出最佳的匹配对象，完成最终的身份认证或相似人群搜索。此时需要衡量特征之间差异大小或关系远近的度量函数，以及将待比对的人脸划分到最佳匹配对象的分类器。

5.4.1　典型的度量方法

根据数据特性的不同，在进行人脸特征比对时需要采用不同的度量方法，欧几里德距离是最常用的度量函数。对于任意的特征向量 a、b 和 c，严格的度量函数满足：

- 非负性：$D(a,b) \geq 0$，当且仅当 $a=b$ 时，$D(a,b)=0$。
- 对称性：$D(a,b)=D(b,a)$。
- 三角不等式：$D(a,b)+D(b,c) \geq D(a,c)$。

常用的度量函数可以分为两类：距离度量和相似性度量。

1. 距离度量方法

距离度量方法将提取到的人脸特征看作高维欧氏空间中的特征点，每个人脸特征对应一个元素 x_i（$i=1,2,\cdots,D$），D 为特征提取后的特征维数，首先，整幅人脸图像的特征构成一个高维的特征向量 $X=[x_1,x_2,\cdots,x_D] \in R^D$；然后构造一个适当函数，计算两个特征点之间的某种距离，距离越大，特征之间的差异越大。

- 欧氏距离（Euclidean Distance）

欧氏距离即欧几里德距离，是最常用、最易理解的一种距离度量方法。

$$Dist(X,Y)=\sqrt{\sum_{i=1}^{D}(x_i-y_i)^2}$$

其中，$X=[x_1,x_2,\cdots,x_D],Y=[y_1,y_2,\cdots,y_D]$ 为两幅人脸图像对应的特征向量。欧氏距离衡量特征在多维空间上存在的绝对距离，与各个特征点所在的位置坐标直接相关，其值越大说明特征之间的差异越大。欧氏距离体现个体数值特征的绝对差异，用于需要从特征维度的数值大小中体现差异的分析。欧氏距离的缺点是将特征的各个分量的量纲同等看待，需要保证各维度的指标具有相同量纲，两个单位不同的指标使用欧氏距离可能使结果失真。另外没有考虑特征的各个分量之间不同的分布情况，如均值和方差等。

- 标准化欧氏距离（Standardized Euclidean Distance）

标准化欧氏距离即使特征的各维度分量的量纲或分布不相同，但是各分量的标准化变量都是均值为 0、方差为 1，具有相同的量纲和分布，其定义为：

$$Dist(X,Y)=\sqrt{\sum_{i=1}^{D}\frac{(x_i-y_i)^2}{\sigma_i^2}}$$

其中，σ_i 为特征的第 i 个分量的标准差。

标准化欧氏距离相当于在进行欧氏距离计算之前，先对数据进行归一化处理。

❑ 马氏距离（Mahalanobis Distance）

记 N 个特征 $X_1, X_2, \cdots, X_N \in R^D$ 之间的协方差为 Σ，特征 X_i 与特征 X_j 之间的马氏距离为：

$$Dist(X,Y) = \sqrt{\left(X_i - X_j\right)^T \Sigma^{-1} \left(X_i - X_j\right)}$$

当协方差矩阵 Σ 为单位矩阵时，即特征的各分量之间独立同分布，马氏距离退化为欧氏距离；当协方差矩阵为对角矩阵时，马氏距离退化为标准化欧氏距离。

马氏距离不受量纲的影响，两点之间的马氏距离与原始数据的测量单位无关；排除分量之间的相关性干扰；考虑各种人脸特征之间的联系，如脸的长度与鼻子的长度是有一定关联的。

协方差矩阵 Σ 的引入会夸大变化微小的分量的作用；要求特征总数 N 大于特征维数 D，否则 Σ 的逆矩阵不存在，这在人脸搜索中难以满足，即“小样本”带来困惑。

2. 相似性度量方法

相似性度量（Similarity）通过计算特征之间的相似程度来度量特征之间的差异。与距离度量方法相反，相似性度量的值越小说明特征之间的相似性越小，差异越大。

❑ 余弦相似性（Cosine Similarity）

余弦相似性是最常见的相似性度量，直接将特征 X 看作高维空间中的向量，两个不同的人脸特征之间的相似性通过两者对应的特征向量的夹角余弦值来度量，其定义为：

$$Sim(X,Y) = \cos\theta = \frac{X \cdot Y}{\|X\|\|Y\|} = \frac{\sum_{i=1}^{D} x_i y_i}{\sqrt{\sum_{i=1}^{D} x_i^2}\sqrt{\sum_{i=1}^{D} y_i^2}}$$

欧氏距离重视距离或长度上的差异，余弦相似性注重两个特征在方向上的差异。

❑ 相关系数（Correlation Coefficient）

相关系数即皮尔逊相关系数（Pearson correlation），衡量两个人脸特征之间的相关程度，其定义为：

$$\begin{aligned} Cor(X,Y) &= Sim(X\text{-}\bar{X}, Y\text{-}\bar{Y}) \\ &= \frac{\sum_{i=1}^{D}\left(x_i\text{-}\bar{X}\right)\left(y_i\text{-}\bar{Y}\right)}{\sqrt{\sum_{i=1}^{D}\left(x_i\text{-}\bar{X}\right)^2}\sqrt{\sum_{i=1}^{D}\left(y_i\text{-}\bar{Y}\right)^2}} \end{aligned}$$

其中，$\bar{X}=\sum_{i=1}^{D}x_i$ 是特征 X 的均值。相关系数具有平移不变性和尺度不变性，常常用于相关性分析之中。其取值范围是[-1,1]，值为正时表示正相关，值为负时表示负相关，值为零时表示两者之间不存在线性相关。两个特征的相关系数绝对值越大，它们之间的相关度越高。

- 杰卡德相似（Jaccard Similarity）

当将提取到的人脸图像的特征看作一个集合时，如 1 表示包含某个特征，而 0 表示不包含该特征，采用杰卡德相似系数来度量两幅人脸图像之间的相似性。其定义为：

$$Jac(X,Y)=\frac{|X\cap Y|}{|X\cup Y|}$$

杰卡德相似系数是两个集合的交集元素在并集中所占的比例。

5.4.2　典型的分类器

分类器对每一个待分类或待搜索的人脸图像赋予一个类别名称或推荐一个匹配对象，常用的有 KNN 分类器、SVM 分类器等。

1. KNN 分类器

KNN（K-Nearest Neighbor，K 最近邻）分类器是一种理论成熟、原理简单的统计分类器，其基本思想是“物以类聚、人以群分”、“近朱者赤、近墨者黑”，由 A 的邻居来推断 A 的类别。该方法首先计算待分类或待搜索的人脸图像与数据库中已经正确分类的人脸图像之间的某种度量，找到和新样本距离最近或最相似的 K 个近邻样本，然后统计这些近邻样本的类别属性，来判定新样本的类别。相关论文为美国 Stanford University 的 Cover 和 Hart 发表于 1967 年的 *Nearest neighbor pattern classification*（*IEEE Transactions on Information Theory*），被引用次数已经超过 5200 次。

具体地说，假设数据库中共有 k 个类别的 N 个人脸图像样本 $x_1^1,\cdots,x_{N_1}^1,\cdots,x_1^k,\cdots,x_{N_k}^k$，其中 x_j^i 是来自第 w_i 个类别的第 j 个样本，$N=\sum_{i=1}^{k}N_i$。对一个待判定类别的新样本 x，它的 K 个最近邻样本中属于每个类别 $w_i, i=1,2,\cdots,k$ 的样本数分别为 $K_1,K_2,\cdots,K_k$，则新样本的类别判决规则为：

$$x\in w_m,\ if\ m=\max_{i=1,2,\cdots,k}K_i$$

即如果新样本的 K 个近邻样本都属于同一类别，则新样本也属于该类别；否则，对候选类别根据“少数服从多数”的投票规则确定新样本的类别。在 KNN 分类器中，K

值的设定对分类的影响较大，K 值太小容易受噪声的影响，K 值太大则可能包含太多其他类别的样本，K 值的设定一般低于数据中已知类别样本的平方根，通过采用交叉检验（cross-validation）来确定。当 K=1 时，就是人脸识别中常用的最近邻分类；对二维平面分类问题，K 通常取奇数以避免投票时正负两类得票相同。

KNN 分类器是一种基于直觉的简单分类器，易于理解和实现，也无需估计参数和训练，在一定程度上还可以降低噪声样本对分类的干扰。在测试各种分类器时，KNN 分类器常常被当成一个基准（baseline）分类器，以便和其他更复杂的分类器进行性能对比。

当已知类别的样本不平衡时，如某个类的样本数很多而其他类的样本数很少，容易导致新样本的 K 个近邻样本中大容量类的样本数占多数，使得新样本往容量大的类别“聚集”。可以通过对不同距离的近邻样本赋予不同权值的方式加以改进，如与新样本距离越近的样本，权值越大，权值为距离平方的倒数。由于需要计算每个待分类样本与全体已知类别样本的度量函数以进行评分，使得 KNN 分类器的计算量较大，内存开销大，不适用于大容量数据库的人脸搜索，可以通过事先采用浓缩技术或编辑技术，去除对分类作用不大的样本来加以改进。

2. SVM 分类器

SVM（Support Vector Machine，支持向量机）分类器是一种监督式学习的分类器，属于线性分类器。SVM 分类器能够在最小化经验误差的同时最大化几何间隔，称为最大间隔分类器（Maximum Margin Classifier）。SVM 分类器的基本思想是：当不同类别之间的分隔间隔越大时，不同类别的人脸样本点分得越开，分类器的总误差越小。

假设人脸特征点用 $x \in R^D$ 来表示，这是一个 D 维向量。在线性可分的情况下，在 D 维的特征空间中存在一个 D-1 维的超平面可以把数据分割开来。

SVM 的目标是最大化几何间隔，实际就是寻找超平面，使得超平面到正/负类样本中最近的点都最远，从而实现分隔的间隙越大越好，把两个类别的点分得越开越好。要实现超平面到正/负类样本中最近的点都最远，容易知道超平面到正/负类样本中最近的点应该是等间距的。

为消除 w,b 的尺度变化对超平面 $w^T x + b = 0$ 的影响，不妨设最近的点满足 $y_i(w^T x_i + b) = 1$，满足该条件的点即是 SVM 中的 Support Vector（支持向量或支持点），得到 SVM 的优化目标函数：

$$\max \frac{2}{\|w\|}$$
$$s.t.\ y_i(w^T x_i + b) \geq 1, i = 1, 2, \cdots, N$$

其中，$2/\|w\|$ 是两个相互平行的支持平面之间的距离，$y_i(w^T x_i + b) \geq 1$，$i = 1,2,\cdots,N$ 表示所有样本点均需大于最近距离。

上式可以等价为，SVM 的原问题：

$$\min \frac{1}{2}\|w\|^2$$
$$s.t.\ y_i(w^T x_i + b) \geq 1, i = 1,2,\cdots,N$$

这使得后续操作（如函数求导等）更容易，并且是一个带约束的二次规划问题（Quadratic Programming，QP），是一个凸优化问题，具有全局最优解。

这个带约束的优化问题可以用拉格朗日乘子法转化为无约束的优化问题，通过一些系数把约束条件和目标函数结合在一起，其拉格朗日目标函数为

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N}\alpha_i(y_i(w^T x_i + b) - 1)$$

为求解该目标函数，首先求解 $L(w,b,\alpha)$ 关于 w,b 的最优解，为此，分别令 $L(w,b,\alpha)$ 关于 w,b 的偏导数等于 0，有：

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{N}\alpha_i y_i x_i$$
$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i y_i = 0$$

将上述表达式代回 $L(w,b,\alpha)$，得到对偶问题：

$$\max_{\alpha} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j$$
$$s.t., \alpha_i \geq 0, i = 1,2,\cdots,N$$
$$\sum_{i=1}^{N}\alpha_i y_i = 0$$

这就是线性可分情况下需要最终优化的式子。对该式的求解，可以采用 SMO（Sequential Minimal Optimization，序列最小优化）方法等。得到 $\alpha_i, i = 1,2,\cdots,N$ 之后，即可对新增数据点 x 进行分类，将 $w = \sum_{i=1}^{N}\alpha_i y_i x_i$ 代入分类函数 $f(x) = w^T x + b$，有

$$f(x) = w^T x + b = \sum_{i=1}^{N}\alpha_i y_i (x_i^T x) + b$$

当 $f(x) < 0$ 时，$y = -1$；当 $f(x) > 0$ 时，$y = +1$。对于新增数据点 x 的分类，只需计算它与训练样本点的内积即可。

5.5　“大海捞针”人脸搜索系统

“大海捞针”人脸搜索系统由国防科技大学 VAP 研究中心设计，采用视觉机器学习方法，对监控视频进行自动分析，根据某人的人脸照片、画像、监控人像、目击者描述等，快速搜索人脸目标，提取图纹特征，与模板目标进行鲁棒比对，实现高效搜索。

5.5.1　体系结构

该系统的搜索流程如图 5.13 所示，首先采用聚类分析和 Adaboost 方法检测人脸；然后针对人脸局部区域设计并训练多个对应的深度神经网络（DNN），用于识别不同类型的人脸局部图像并计算其显著度；最后，依据显著度水平动态综合人脸局部比对结果，形成最终的分类，完成人脸搜索。

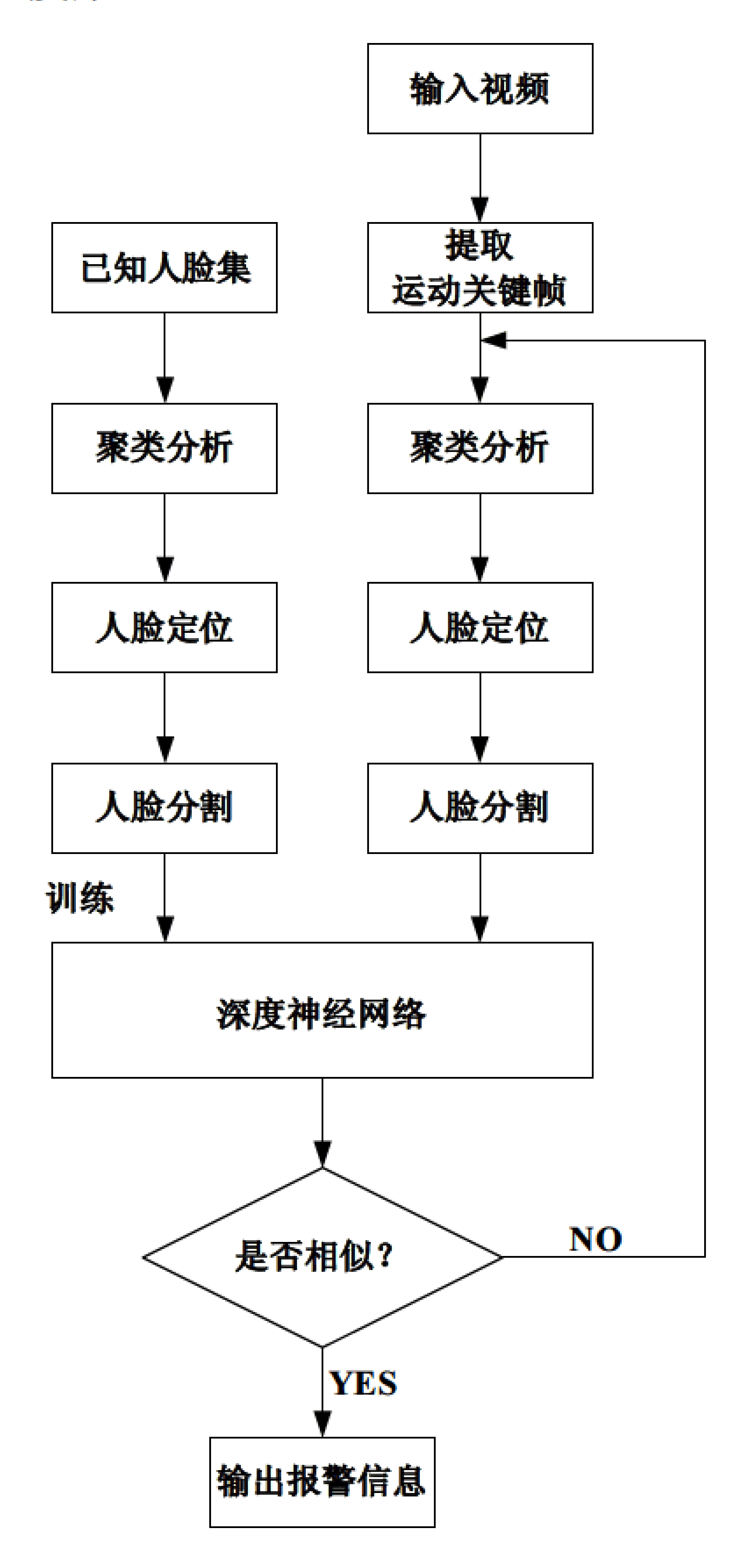


图 5.13　“大海捞针”人脸搜索系统流程

5.5.2　关键技术

大规模人脸搜索系统应重点考虑处理效率和质量，关键技术包括以下内容。

1. 基于聚类分析和 Adaboost 的人脸快速准确检测

根据人脸区域内灰度变化缓慢的特点，首先对图像灰度进行聚类分析，提取出灰度相近的若干区域；然后针对此类区域采用 Adaboost 方法检测和定位人脸。

2. 基于深度神经网络的人脸特征提取

为体现人脸局部特异性，将人脸分割为若干局部区域，对每个区域构造对应的深度神经网络进行特征提取。针对产生的多个 DNN 结果综合问题，对各个 DNN 输出结果进行动态加权综合，使人脸的局部特异性特征得到体现。

3. 基于深度神经网络的人脸特异性比对

在建立多个人脸局部 DNN 的基础上，将目标人脸和模板人脸部件分别输入 DNN，输出两种人脸各局部部件分类的类型及权重，综合生成两种人脸的整体相似矢量，从而可计算其最终相似度，实现人脸搜索。

5.5.3　算法伪代码

大规模人脸搜索系统的关键算法包括：基于 K 均值的人脸聚类、基于 Adaboost 的人脸检测、基于 DNN 的人脸特征提取和分类。

算法 5.1　基于 K 均值的人脸聚类

输入：当前帧像素点 x，总的聚类类别 k。

过程：1. 初始化：

随机选择初始聚类中心 $C_1^{m1}, C_2^{m2}, C_3^{m3}, \cdots, C_k^{mk}$。

其中 $m1, m2, m3, \cdots, mk$ 为各自迭代运算次数。

2. 迭代运算：

while $C_i^{r+1} \neq C_i^r, (i=1,2,3,\cdots,k)$ do
 for $i=1:k$
  if $d(x,C_i^r)=\min(d(x,C_j^r), j=1,2,3,\cdots,k)$
   $S_i^r \leftarrow x$
   $C_i^{r+1}=\frac{1}{n_i^r}\sum_{x\in C_i^r} x, (i=1,2,3,\cdots,k)$
   n_i^r 为属于类别 S_i^r 的像素点数目。
  end if
 end for
end while

输出：当前帧聚类结果 $\{S_1,S_2,\cdots,S_k\}$。

算法 5.2 基于 Adaboost 的人脸检测

输入：$\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$ 为训练样本集，其中 $x_i, i=1,\cdots,N$ 为训练样本，$y_i \in \{0,1\}$ 为样本标签；K 为最大迭代次数。

过程：1. 初始化：

对标签为 $y_i=1$ 的样本，初始化其权值为 $w_i^1=1/2l$；而对 $y_i=0$ 的样本，$w_i^1=1/2m$；其中 l,m 为相应标签样本的总数 $l+m=N$。

2. 进行迭代运算：

for k=1:K

归一化权值，$w_i^k=\frac{w_i^k}{\sum_{i=1}^N w_i^k}$；

对每个 Haar-like 特征，生成对应的弱分类器 $h_j(x), j=1,\cdots,M$，其中 M 为矩形特征的总数；

计算相对于当前样本权值的误差 $\varepsilon_j^k=\sum_{i=1}^N w_i^k\left|h_j(x_i)-y_i\right|$，选取对应最小误差值 $\varepsilon^k=\arg\min_j \varepsilon_j^k$ 的弱分类器 $h_k(x)$ 加入强分类器中；

更新样本权值，$w_i^{k+1}=w_i^k\left(\frac{\varepsilon^k}{1-\varepsilon^k}\right)^{1-e_i}$，若样本 x_i 被 $h_k(x)$ 正确分类，则 $e_i=0$；

否则 $e_i = 1$。

End for

输出：最终的强分类器

$$H(x)=\begin{cases}1, & 如果\sum_{k=1}^{K}\alpha_k h_k(x)\geq\sum_{k=1}^{K}\alpha_k/2\\0, & 其他\end{cases}$$

其中 $\alpha_k=\ln\left(1-\varepsilon^k\right)-\ln\varepsilon^k$。

算法 5.3　基于 DNN 的人脸特征提取和分类

输入：已知分类的训练图像集 $\{I,Y\}$，其中 I 为输入图像数据，Y 为对应分类。

过程：1. 初始化：

初始化：以分布 $U(a^{-0.5},a^{-0.5})$ 随机初始化权重矢量 W_{ij}^{n}，$b_j^n=0$，其中 $a=\max(|Y_j^{n-1}|,|Y_j^n|)$。

2. 迭代运算：

while 分类正确率 $\leqslant C_t$ do

for $i=1:l$

$h^0\left(x_t\right)\leftarrow x_t$

for $j\in\{1,\cdots,i-1\}$

$a^j\left(x_t\right)=b^f+W^j h^{j-1}\left(x_t\right)$

$h^j\left(x_t\right)=sigm\left(a^j\left(x_t\right)\right)$

end for

$o\left(x_t\right)=h^{l+1}\left(x_t\right)=softmax\left(a^{l+1}\left(x_t\right)\right)$

end for

计算当前的分类正确率

$$b^{l+1}\leftarrow b^{l+1}+\varepsilon_{fine-tune}\frac{\partial\log o_{y_t}\left(x_t\right)}{\partial a^{l+1}\left(x_t\right)}$$

$$W^{l+1}\leftarrow W^{l+1}+\varepsilon_{fine-tune}\frac{\partial\log o_{y_t}\left(x_t\right)}{\partial a^{l+1}\left(x_t\right)}h^l\left(x_t\right)^T$$

end while

输出：DNN 网络内部各层权重和偏移量。

5.5.4 性能评价

大规模人脸搜索系统性能评价的常用准则是有效性、效率和灵活性，搜索的有效性，即搜索结果的正确与否最重要。

有效性评价包括使用者的主观感受、量化的评价标准，主观感受易受个体影响，客观评判标准主要有查准率和查全率。

如图 5.14 所示，Q 为人脸图像数据库，A 代表相关图像的集合，B 代表搜索出的人脸图像集合。图中 a+b+c+d=Q，a+c=A，a+b=B。

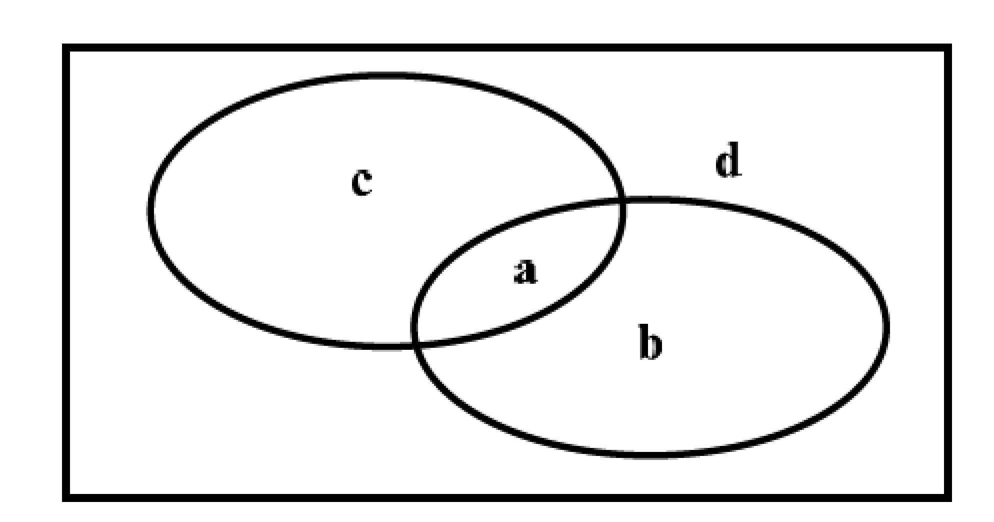


图 5.14 查全率和查准率的关系

1. 查准率

查准率表示一次搜索过程中，系统返回的相似人脸图像个数占所有返回人脸图像个数的比例。正确的主观相似人脸图像越多，查准率越高。

$$P = p(A|B) = \frac{p(A \cup B)}{p(B)} = \frac{a}{a+b}$$

2. 查全率

查全率表示一次搜索过程中，系统返回的搜索结果中相似人脸图像个数占图像库中所有主观相似人脸图像个数的比例。

$$R = p(B|A) = \frac{p(A \cup B)}{p(A)} = \frac{a}{a+c}$$

用查全率 R 作为 x 轴，查准率 P 作为 y 轴，绘制查准率-查全率曲线，即 PVR 曲线。设 PVR 曲线为 $f(x,y)$，则 $f(x,y)$与坐标轴围成的面积为：

$$S_f = \int_0^1 f(x,y)dx$$

S_f为 PVR 指数，该指数越大，搜索性能越好。查全率反映搜索的全面性，查准率反映搜索的准确性。

5.5.5　系统搜索效果

搜索效果如图 5.15 所示。



图 5.15　人像搜索效果

第 6 章

高清卡口车辆信息搜索系统

随着我国社会经济的不断发展，汽车拥有率不断上升，交通发展迅速，车辆管理难度越来越大，交通拥挤、交通事故、违章逃逸、汽车盗窃等发生率显著上升，高清卡口车辆信息搜索系统可以有效促进车辆管理、流量控制、高速公路收费登记和车辆身份认证等。

6.1 车辆信息搜索

高清卡口车辆信息搜索是计算机视觉与模式识别在交通领域中的重要应用，用于对交通卡口的高清摄像视频进行分析，自动识别车牌和车标等信息。

高清卡口车辆信息搜索主要应用于以下 3 方面。

1. 交通

- 闯红灯识别系统

自动抓拍并识别闯红灯的违章车辆号牌信息，将该车违章行为记录在案，作为处罚依据，起到规范行车及警示作用。

- 超速报警系统

当车速超过一定数值时，捕获超速的违章车辆图像，识别其车牌号码并记录，将车牌信息上传至管理部门。

2. 公安

- 嫌疑车辆稽查

高清卡口摄像机不间断采集路面车辆图像，相关设备对采集图像进行分析，识别其中的车牌与车标信息，与通缉、挂失、肇事逃逸、涉案车辆等黑名单比对，一旦相符立即报警。

3. 是收费站

- 高速公路收费系统

车辆进出高清卡口均进行车牌、车标搜索，进站与出站的车辆信息必须一致，有效防止倒卡、换卡等偷逃过路费的行为，阻止中途互换入口卡的逃费车辆，减少车辆停靠时间，加快通行速度。

- 停车场收费系统

进站与出站的车牌、车标必须一致，解决因人员作弊造成的款项流失问题，降低车辆被盗风险，减少工作人员劳动强度。

高清卡口车辆信息搜索可以加强公路、道路管理，减少交通事故、预防车辆被盗案件，可提供全方位、多方式、高效可靠的实用服务，具有广阔的应用市场。

6.2　车牌搜索子系统

6.2.1　车牌搜索概述

我国标准汽车牌照具有如下特点：

- 悬挂位置不统一；
- 由汉字、英文字母和阿拉伯数字组成；
- 根据车型、用途等规定多种格式；
- 底色和字符颜色有多种组合，如蓝底白字、黄底黑字等。

如图 6.1 所示，车牌搜索分析视频图像文件，首先检测并定位可能存在的车牌区域；然后对车牌字符进行分割；最后分类车牌字符。

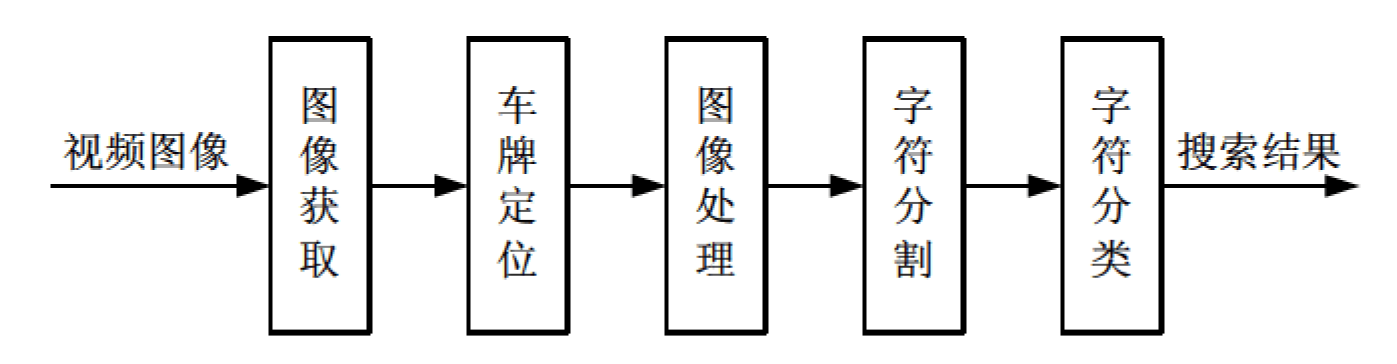


图 6.1　车牌搜索流程

车牌检测与搜索具有如下困难。

1．一步车牌定位与字符分割问题

在图像中确定车牌位置，提取出车牌图像，分割出车牌字符。由于车牌图像采集时受到雨天、大雾、强光等因素影响，车牌图像质量出现不同程度的差异，车牌位置不固定，车牌大小各异，给车牌定位和字符分割带来困难。

2. 高清图像与搜索速度的矛盾

高清图像覆盖面广，可能会同时出现多个车牌；高清视频码流大，需要资源多，分析速度慢，可能导致出现漏车现象，难以实现车辆抓拍率和车牌搜索准确率的提升。

3. 对污损车牌的搜索效果不好

在应用环境中，车牌难免出现污染和磨损现象，如何提高车牌搜索的识别能力是实际需要解决的难题。

6.2.2　车牌区域定位

在视频图像中，根据车牌区域特征判断是否存在车牌图像，若存在则将车牌区域从图像中分割出来。如图 6.2 所示为我国常规车牌细节。



图 6.2　我国常规车牌细节

从机器视觉角度出发，我国车牌具有颜色、形状、投影等直观特征：

- 车牌颜色通常与车身背景、字符颜色等有较大差异；
- 车牌具有连续的矩形轮廓边框，该轮廓常因磨损而不连续；

- 车牌区域内具有若干个基本呈水平排列的字符，字符存在丰富的边缘，呈现较明显的纹理特征；
- 车牌区域内字符之间的间隔均匀，字符和牌照底色各自具有均匀的灰度；
- 属于同一国家或地区的车牌，其长宽比基本固定。

1. 基于颜色特征的车牌区域检测

以我国大陆为例，现有车牌主要包括 4 种颜色类型：蓝底白字为小功率汽车牌照、黄底黑字为大功率汽车牌照、白底黑字/红字为军警用车牌、黑底白字为国外驻华机构所用车牌，车牌底色共有蓝、黄、白、黑 4 种颜色。通过对大量的真实车牌颜色进行分析，可得出 4 种色彩所限定的区间范围，如表 6.1 所示。

表 6.1　车牌区域 HSV 特征表（“/”表示无用信息）

	蓝	黄	白	黑
H	[190~245]	[25~55]	/	/
S	[0.35~1]	[0.35~1]	[0.35~0.1]	/
V	[0.3~1]	[0.3~1]	[0.91~1]	[0~0.35]

基于颜色特征的车牌区域检测方法简便、直观。在 HSV 空间中，在 V 分量上设定区间范围可以将黑色区域识别出来，将其灰度值设为 100；类似地，综合 H、S 分量可区分蓝、黄区域，将蓝色和黄色区域的灰度值分别设为 255 和 200；综合 V、S 分量可以识别白色区域，将其灰度值设为 150；将其他颜色信息设为背景，灰度值设为 0。经过上述处理，原始视频图像被转化为 5 级灰度图，可以快速定位到与车牌颜色有关的区域。基于车牌的 4 种颜色特征，在视频图像中进行搜索，可初步确定可能存在车牌的区域。

算法 6.1　车牌检测颜色模型

输入：监控视频图像帧。
过程：1. 提取图像 RGB 值；
　　2. 将 RGB 值转换为 HSV 值；
　　3. 依据帧图像各像素 HSV 值、车牌颜色范围进行检测。
输出：可能的车牌区域。

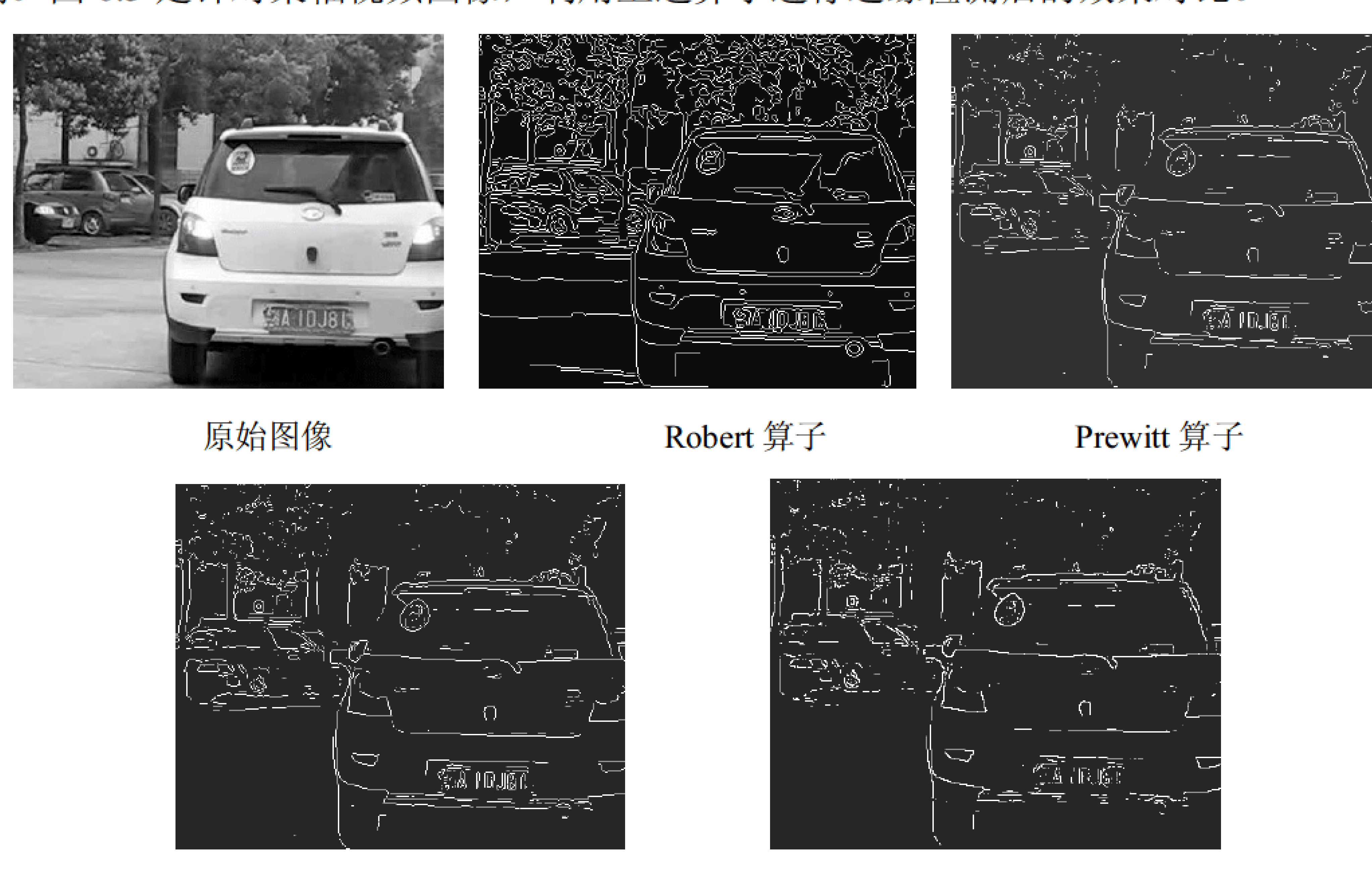
基于颜色特征的车牌定位方法主要依赖于车牌区域的颜色属性，原理简单，实现快速。在实际应用中，可能存在视频降质导致色差，以及当车身、环境与车牌颜色相近时，可造成车牌区域检测的漏定位或者错定位。

2. 基于形状特征的车牌区域检测

在待处理视频图像中，搜索车牌区域固有的几何形状特征，如边缘特征、整体轮廓特征、局部矩形连通区域等，发现可能存在的车牌区域。

（1）边缘定位方法

数字图像中边缘的特点包括：其两侧分属于两个区域，各区域内部灰度相对均匀一致，而这两个区域之间的灰度存在较大差异，交界处形成边缘。边缘检测的目的是在抑制噪声的前提下精确定位边缘。检测的边缘算子有多种，如 Roberts 算子、Prewitt 算子、Sobel 算子、Laplace 算子等。上述算子利用物体边缘处灰度变化相对剧烈的特点，可以检测图像中可能存在的边缘。各算子对不同边缘类型的敏感程度不同，检测结果也有差别。图 6.3 是针对某幅视频图像，利用上述算子进行边缘检测后的效果对比。



原始图像　　Robert 算子　　Prewitt 算子

Sobel 算子　　拉普拉斯算子

图 6.3　边缘检测

Robert 算子利用局部方差寻找图像边缘，检测效果比较精确；Prewitt 算子和 Sobel 算子对噪声具有一定的抑制能力，但不能完全排除噪声影响；拉普拉斯算子采用二阶微分算子，对图像中的阶跃型边缘点检测准确且检测结果具有旋转不变性，但该算子容易丢失部分边缘的方向信息，同时抗噪能力较差。针对不同的环境和要求，应合理选择恰当的算子用于边缘检测，才能达到更好的效果。当检测到边缘之后，再具体研究各边缘

之间的方向、位置关系，当搜索到大致围成矩形框的四条边缘时，则可初步定位该四条边缘围成的矩形区域即为车牌区域。定位流程如图 6.4 所示：

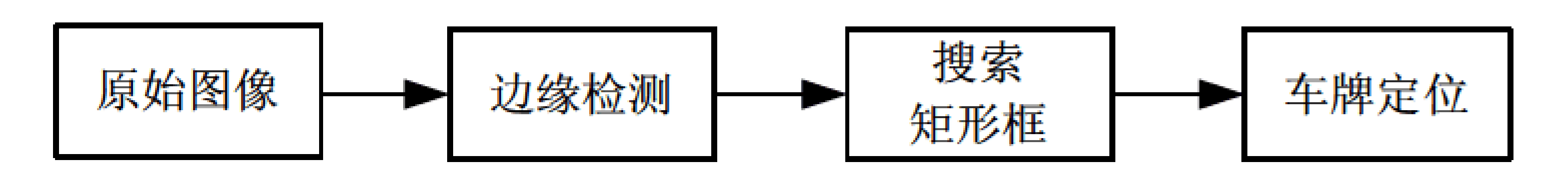


图 6.4 基于边缘检测的车牌定位

边缘定位方法的定位准确率较高，能有效降低噪声干扰，可适用于存在多个车牌的视频图像。但是该方法耗时较长，并且对车牌褪色和图像降质的情况，由于检测不到牌照边缘会导致定位失败；当存在外界干扰以及车牌倾斜时，定位后的区域比车牌稍大。

（2）模板匹配方法

在实际应用中，摄像机高度和角度确定后，获取的图像就相对稳定，车牌的大小变化范围较小。因此可以定义一个尺寸略大于实际图像中待处理牌照大小的模板，并用该模板对整个图像逐点扫描，统计各个模板区域内边缘点的个数。如果某一区域内的边缘点个数达到一定的比例，就认为该区域是一个牌照的候选区域。由于对整幅图进行搜索耗时较长，为了加快搜索速度，可采用分块策略。具体步骤如下。

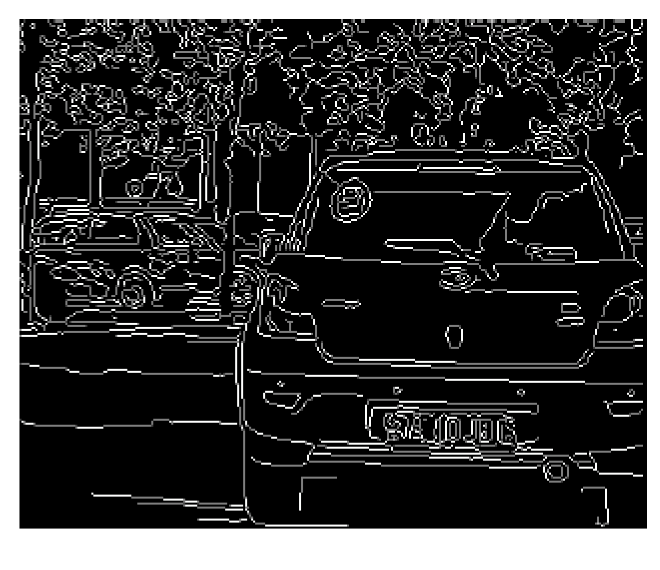
步骤01 假设实际图像中车牌长宽统计信息为 $m \times n$ 像素，将模板预设为 A，尺寸为 $m/8 \times n/8$ 像素，并将模板内的值初始化为 1；

步骤02 假设待处理视频图像分辨率为 $M \times N$ 像素，将其分成 8×8 的块，计算每块内的边缘点个数，将其存入一个矩阵 B 中，B 为能量块矩阵，维数为 $M/8 \times N/8$；

步骤03 用模板 A 在 B 中逐像素点进行卷积运算，计算 B 中每个点对应的值，并将其存入矩阵 $\boldsymbol{C}$ 中。此值越大，则原图对应区域边缘点个数越多，判断其为候选牌照区域。

在实际应用中，对于得到的 M/h 个候选牌照区域，需按照矩阵 $\boldsymbol{C}$ 值的大小和位置信息排序：将排序靠前的几个作为最终的车牌候选区域。排序方法可以先按照矩阵 $\boldsymbol{C}$ 值的大小排序，得到几个最大可能的牌照区域，然后再按照位置信息进行排序；也可以将位置和 $\boldsymbol{C}$ 中的值加权平均处理后排序。

在牌照区域的同行高度上以及相邻的上下区域干扰信息较少，牌照区域会落在 M/h 个候选区域之中，偏下方的候选区域为牌照区域的可能性较大。图 6.5 为按照模板匹配方法进行车牌定位的效果图。









边缘图像　　　提取到三个候选区域　　　最终确定的车牌区域

图 6.5　模板匹配车牌定位方法

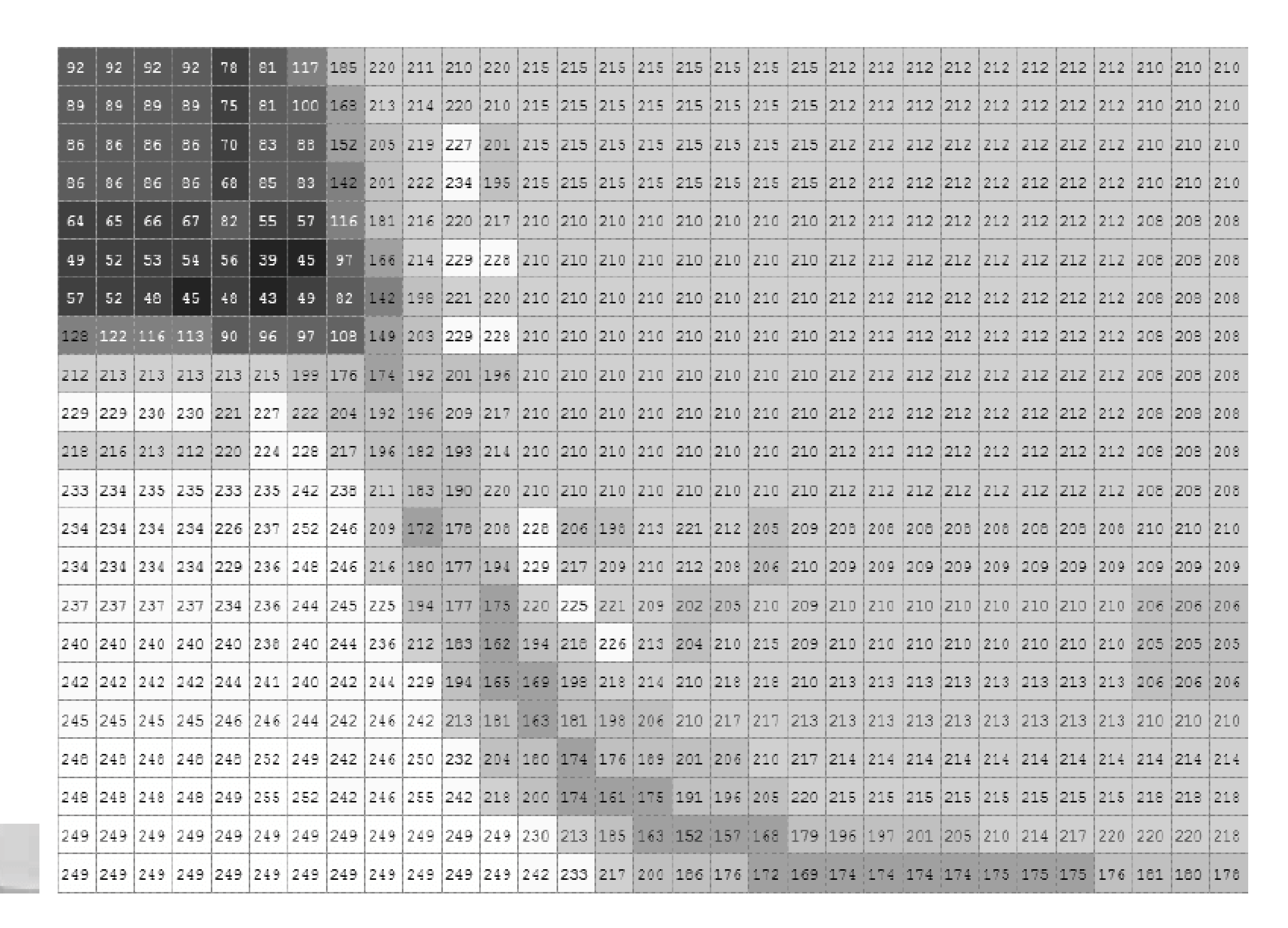
（3）纹理特征方法

纹理反映物体表面颜色或灰度的某种变化，与物体本身属性相关，纹理特征可直观地描述区域的平滑、稀疏、规则性等特性。

我国车辆边缘在灰度上呈现屋顶状边缘。在车牌区域内部，字符和牌底的灰度均匀地呈现波峰波谷，形成比较稳定的纹理特征。图 6.6 显示的是车牌区域与非车牌区域图像灰度的差异。

62	62	62	62	62	98	95	92	87	82	77	74	72	70	95	127	139	130	109	90	81	92	94	100	103	105	105	102
89	85	92	100	102	131	123	76	68	53	35	62	61	67	76	97	112	108	89	85	93	81	85	77	81	70	77	100
89	80	84	102	122	186	163	100	75	57	47	78	79	101	90	81	76	70	77	107	140	135	103	75	84	75	75	120
94	81	78	106	143	217	187	119	85	70	70	94	93	99	90	81	76	70	79	116	155	175	112	71	94	85	73	133
100	89	81	111	159	205	179	127	93	85	93	101	93	82	83	90	94	87	90	117	153	168	109	72	101	93	77	131
102	89	84	122	176	201	179	139	94	83	98	99	94	91	91	92	90	82	89	127	169	157	114	81	97	89	88	139
98	80	88	136	192	211	192	158	95	73	94	97	102	96	97	100	96	83	89	130	175	163	127	88	86	80	97	157
96	78	96	149	196	190	184	172	103	72	96	93	105	91	101	111	109	92	91	126	167	166	130	85	83	77	97	160
97	84	106	152	188	145	158	172	112	81	101	91	100	102	107	112	106	86	88	127	174	159	117	81	89	82	92	150
105	93	118	165	183	102	123	178	146	97	102	107	133	139	131	101	101	95	93	136	161	149	112	83	97	92	99	153
89	88	139	182	173	103	129	180	144	92	98	105	119	123	119	94	97	94	95	143	171	149	112	83	97	92	99	155
78	93	164	190	149	88	128	177	143	92	98	99	97	106	105	84	91	90	95	148	180	149	111	81	95	92	100	156
84	110	176	185	140	90	136	177	145	97	95	99	82	97	99	82	89	86	91	147	182	145	110	79	95	94	100	157
96	128	178	185	163	142	173	187	152	99	88	105	90	99	103	88	90	82	83	140	177	145	109	78	95	94	102	159
97	138	179	186	185	188	201	191	163	105	84	110	105	102	110	94	94	81	80	136	175	142	107	78	95	94	103	160
91	144	180	168	156	151	166	171	179	133	88	109	101	102	111	98	97	82	81	140	181	141	105	77	94	94	103	161
89	149	181	143	102	74	106	143	195	162	100	102	87	99	110	98	98	83	83	144	188	141	104	77	94	94	104	163
98	122	128	106	79	69	95	126	135	123	106	97	95	95	93	95	95	90	94	122	156	127	111	92	82	94	125	161
106	115	116	108	97	94	94	97	103	111	119	122	123	106	99	102	110	110	106	114	128	109	103	97	95	102	116	132
127	118	113	115	122	130	122	111	106	109	118	130	134	124	111	106	114	117	110	107	113	103	107	114	119	120	118	112
144	129	119	123	132	130	141	148	141	126	122	131	144	155	139	127	127	126	122	128	143	124	128	134	136	136	132	127
138	128	119	114	114	107	125	143	145	136	131	138	147	160	157	157	156	146	138	153	176	146	141	135	131	132	138	146
117	112	105	98	93	113	113	113	118	120	120	118	114	121	131	147	155	141	125	134	153	140	131	119	112	113	124	140

车牌区域图像及灰度特征



92	92	92	92	78	81	117	185	220	211	210	220	215	215	215	215	215	215	215	215	212	212	212	212	212	212	212	212	210	210	210
89	89	89	89	75	81	100	168	213	214	220	210	215	215	215	215	215	215	215	215	212	212	212	212	212	212	212	212	210	210	210
86	86	86	86	70	83	88	152	205	219	227	201	215	215	215	215	215	215	215	215	212	212	212	212	212	212	212	212	210	210	210
86	86	86	86	68	85	83	142	201	222	234	195	215	215	215	215	215	215	215	215	212	212	212	212	212	212	212	212	210	210	210
64	65	66	67	82	55	57	116	181	216	220	217	210	210	210	210	210	210	210	210	212	212	212	212	212	212	212	212	208	208	208
49	52	53	54	56	39	45	97	166	214	229	228	210	210	210	210	210	210	210	210	212	212	212	212	212	212	212	212	208	208	208
57	52	48	45	48	43	49	82	142	198	221	220	210	210	210	210	210	210	210	210	212	212	212	212	212	212	212	212	208	208	208
128	122	116	113	90	96	97	108	149	203	229	228	210	210	210	210	210	210	210	210	212	212	212	212	212	212	212	212	208	208	208
212	213	213	213	213	215	199	176	174	192	201	196	210	210	210	210	210	210	210	210	212	212	212	212	212	212	212	212	208	208	208
229	229	230	230	221	227	222	204	192	196	209	217	210	210	210	210	210	210	210	210	212	212	212	212	212	212	212	212	208	208	208
218	216	213	212	220	224	228	217	196	182	193	214	210	210	210	210	210	210	210	210	212	212	212	212	212	212	212	212	208	208	208
233	234	235	235	233	235	242	238	211	183	190	220	210	210	210	210	210	210	210	210	212	212	212	212	212	212	212	212	208	208	208
234	234	234	234	226	237	252	246	209	172	178	208	228	206	198	213	221	212	205	209	208	208	208	208	208	208	208	208	210	210	210
234	234	234	234	229	236	248	246	216	180	177	194	229	217	209	210	212	208	206	210	209	209	209	209	209	209	209	209	209	209	209
237	237	237	237	234	236	244	245	225	194	177	175	220	225	221	209	202	205	210	209	210	210	210	210	210	210	210	210	206	206	206
240	240	240	240	240	238	240	244	236	212	183	162	194	218	226	213	204	210	215	209	210	210	210	210	210	210	210	210	205	205	205
242	242	242	242	244	241	240	242	244	229	194	165	169	198	218	214	210	218	218	210	213	213	213	213	213	213	213	213	206	206	206
245	245	245	245	246	246	244	242	246	242	213	181	163	181	198	206	210	217	217	213	213	213	213	213	213	213	213	213	210	210	210
248	248	248	248	248	252	249	242	246	250	232	204	180	174	176	189	201	206	210	217	214	214	214	214	214	214	214	214	214	214	214
248	248	248	248	249	255	252	242	246	255	242	218	200	174	161	175	191	196	205	220	215	215	215	215	215	215	215	215	218	218	218
249	249	249	249	249	249	249	249	249	249	249	249	230	213	185	163	152	157	168	179	196	197	201	205	210	214	217	220	220	220	218
249	249	249	249	249	249	249	249	249	249	249	249	242	233	217	200	186	176	172	169	174	174	174	174	175	175	175	176	181	180	178

非车牌区域图像及灰度特征

图 6.6　车牌及非车牌区域图像灰度特征对比

基于灰度纹理特征进行车牌定位的处理流程如图 6.7 所示。

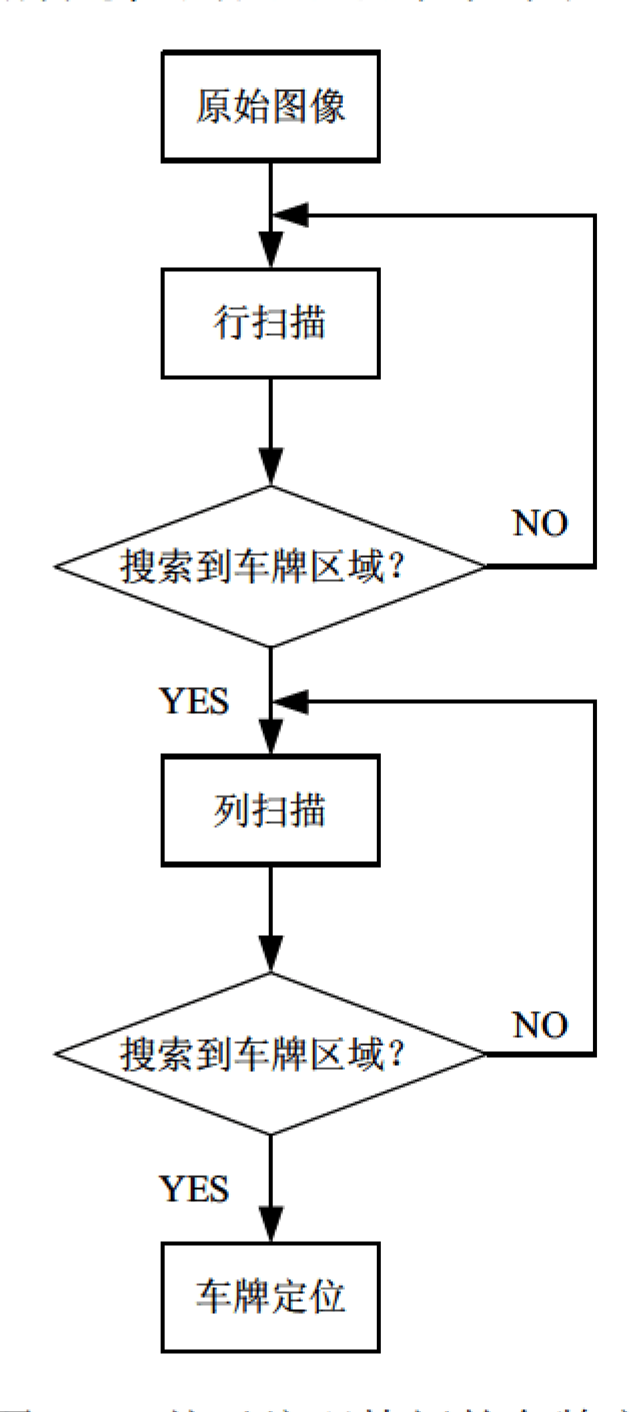


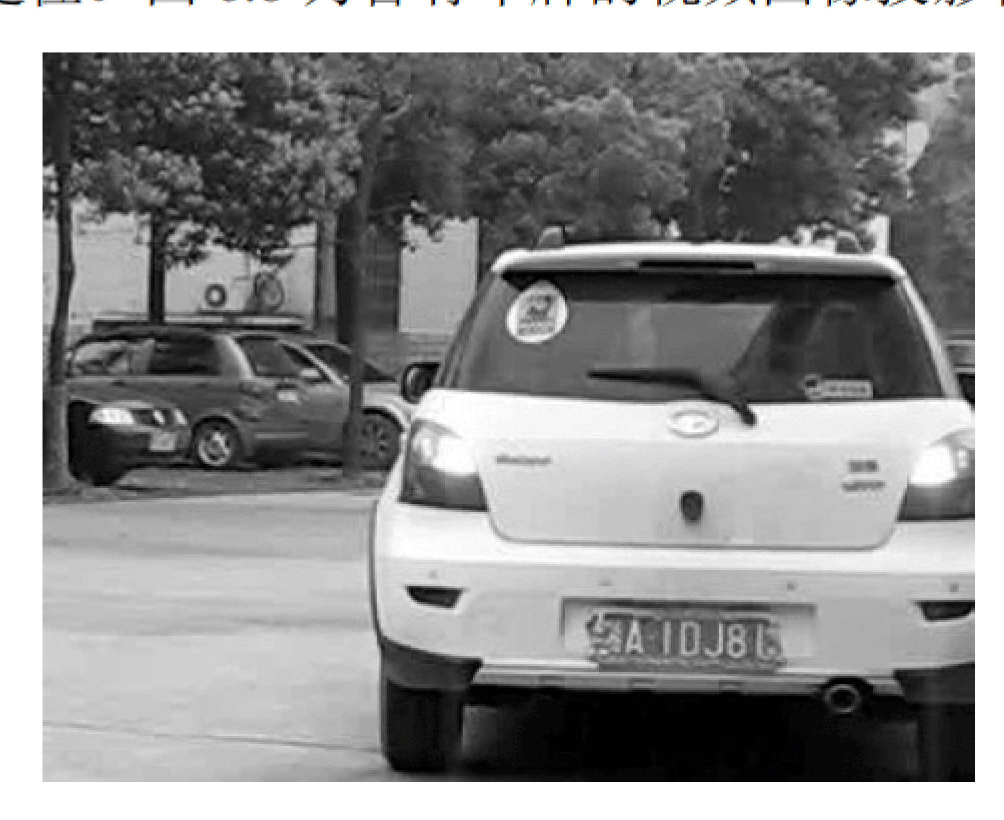
图 6.7　基于纹理特征的车牌定位

在该处理中，基于灰度图像进行行扫描，找出图像中每一行可能的车牌线段，记录它们的起始坐标和长度。如果连续若干行均存在车牌线段，且行数大于某一预设阈值，则可判断为在行方向检测到一个车牌候选区域，并记录该候选区域的起始行和高度。针对已检测到可能存在车牌的区域进行列扫描，获得该车牌候选区域的起始行和高度，结合前一步骤获取的起始列坐标和长度，从而确定一个车牌区域；继续在其他可能存在的车牌区域进行类似搜索，直至遍历完成所有的车牌候选区域。

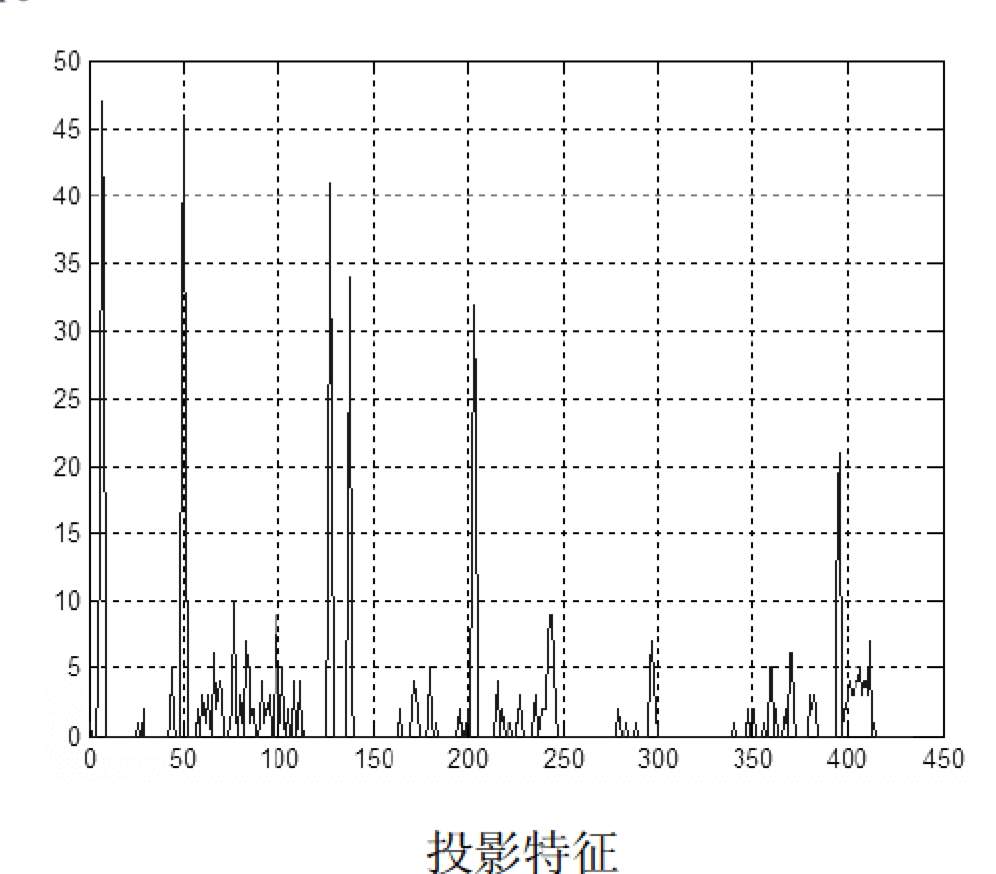
纹理特征方法对于牌照倾斜、变形、光照不均具有较好的适应性，但对噪声比较敏感。针对背景复杂的图像，可以将纹理特征与垂直投影相结合，有效地降低复杂背景的干扰。

3. 基于投影特征的车牌区域检测

车牌区域有丰富密集的边缘信息，通过投影转换可表现出明显的形态特征。牌照区域的水平投影表现为连续的波峰区域，区域内部没有大的落差，波形平缓，与其他小的波峰区域有明显的波谷间隔。牌照区域的垂直投影表现为一组密集的小峰群，各个小波峰区域间距较小，符合波峰合并条件，可以合并成一个大的波峰区域。这组小峰群与其他峰群有较大的间隔，可以明显区分。依据这些投影特征，可以对牌照进行水平和垂直定位。图 6.8 为含有车牌的视频图像投影特征。



原始图像



投影特征

图 6.8　车牌视频图像投影

投影方法主要依据投影图像波峰形态特征进行处理。

首先，在原车牌图像上进行垂直边缘检测，将垂直边缘投影到纵轴上。车牌区域因为垂直边缘密集，因此在对应位置投影存在尖峰，并且在车牌位置以外的区域曲线较平缓，没有明显尖峰。基于该特征，沿由下向上的方向，设置适当阈值搜索投影曲线中的

有效波峰，即可定位出车牌区域的垂直位置。

然后，对提取到的区域进行水平定位。其方法是将垂直定位得到的车牌区域图像与拉普拉斯算子进行卷积运算，然后将该边缘图像进行水平投影，在没有车牌的位置投影曲线值相对较小，且变化平缓，反之有车牌的位置投影曲线值较大，且变化剧烈。该投影图中最大的波峰位置即为对应车牌的左右边界。

6.2.3　车牌字符分割

可将完整车牌图像分割为单个字符图像，以缩小识别范围，提高识别质量。常用的车牌字符分割方法包括基于结构特征、基于形态特征以及基于投影的分割方法。

1. 基于结构特征的分割方法

我国车牌由汉字、数字和字母组成，相对数字和字母而言，汉字具有更明显的结构特征差异，可将车牌字符分为汉字和非汉字两大类。

（1）汉字字符分割

如图 6.9 所示，车牌字符中的汉字结构关系可分为上下、左右与包围关系。

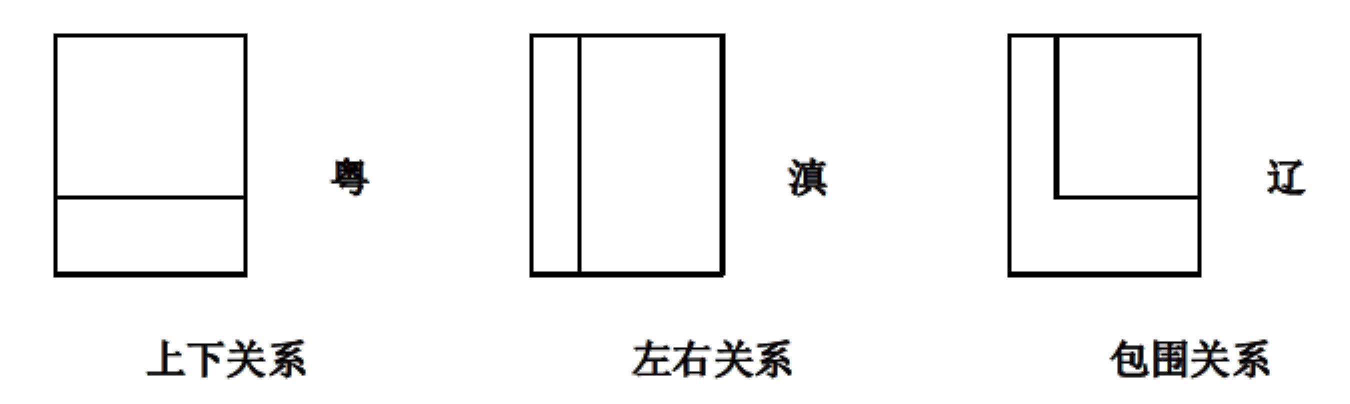


图 6.9　汉字结构关系

将汉字的部件按左右或上下顺序排列，其高度和宽度在整个汉字字符中占主要成分的部件为汉字的主体部件，如“粤”的上部、“辽”的左下部、“滇”的右部为主体部件。

如图 6.10 所示，利用汉字的结构关系与主体部件，可对汉字进行分割。

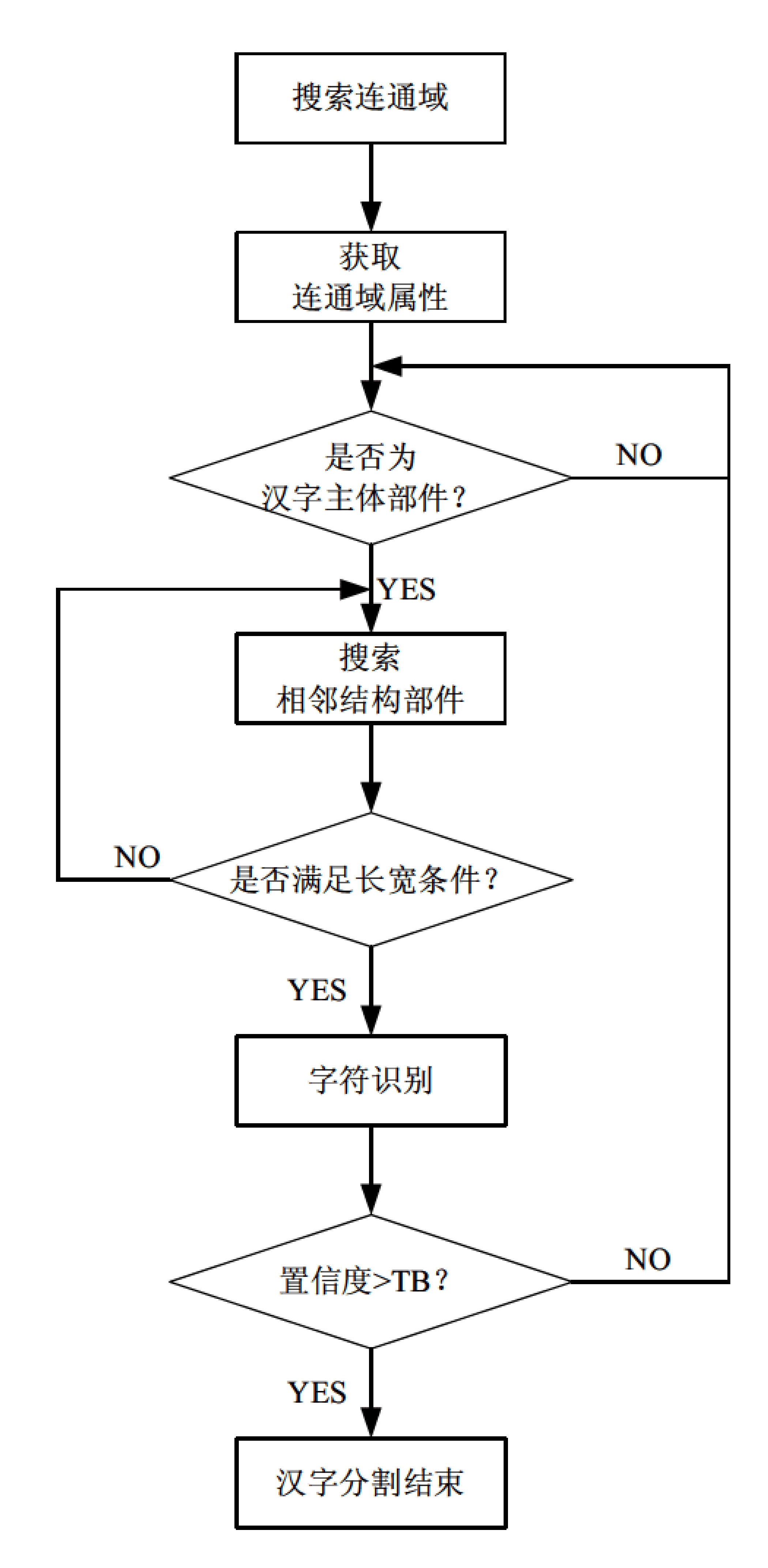


图 6.10　基于结构特征的汉字字符分割

汉字字符的分割步骤如下。

步骤 01　获取车牌连通域属性

在二值车牌图像上，首先搜索并获取车牌图像的连通域信息，记录每个连通域的属性：右边界、左边界、上边界、下边界、像素数目、宽度、高度、水平中心等。

利用连通域的属性和车牌字符的特征确定车牌字符串的以下属性值：

$$CH, CW \approx CH/2 \text{ 和 } CWZ$$

其中，CH 为字符高度，CW 为常规字符（除 1 外）的宽度，CWZ 为连通域的水平

中心距离。

在车牌字符中，数字和字母字符都具有单连通特性，汉字的主体部件具有单连通特性。

步骤 02　选择汉字主体部件

如果某个连通域满足下面两个条件之一，则认为是汉字字符的主体部件。

a、$0.5\times CW<$ 宽度 $<1.2\times CW$，$0.5\times CH<$ 高度 $<1.2\times CH$。

b、$2\times CW/3<$ 宽度 $<1.2\times CW$，$1\times CH/3<$ 高度 $<1.2\times CH$。

步骤 03　确定非主体部件

根据组成汉字部件的结构关系(上下、左右、包围)确定组成汉字的其他部件。步骤如下。

a. 向汉字结构中加入新部件，字符宽度在 $(0.8\times CW,1.2\times CW)$ 之间；

b. 依次搜索与主体部件是左包围、右包围、上下和左右关系的部件。假如合并后的汉字满足条件 a，将其判断为汉字区域。

步骤 04　汉字验证

在搜索完成组成汉字的所有部件之后，即对分割后的汉字进行识别。如果字符识别结果其置信度>*TB*，认为检测到汉字。若置信度<*TB*，则步骤 2 选择的部件不是汉字的主体部件，取下一个可能为主体部件的连通域作为主体部件，回到步骤 3 重复定位汉字。

（2）非汉字字符分割

非汉字字符即英文和数字字符都具有单连通特性，高度在 $(0.8\times CH,1.2\times CH)$ 范围内，与汉字之间的间隔应该在 $(0.8\times CWZ,1.2\times CWZ)$ 范围内，从左至右寻找符合该范围的连通区域进行分割。

在实际处理过程中，非汉字字符区域可能存在以下 3 种情况：

- 没有满足条件的连通域，则判断该区域无字符;
- 仅检测到一个连通域，检查其宽度是否在 $(0.8\times CW,1.2\times CW)$ 范围内，如果满足则该连通域可能为字符区域；否则如果宽度大于 $1.2\times CW$，则对该区域进一步进行分裂，逐个判别是否为字符;
- 具有多个连通域，应按照上述方法对于连通域逐个进行筛选。

基于结构特征的分割方法利用结构特征对汉字进行分割，而针对数字和字母部分，主要依据连通域特性实现分割。该方法实现简单，速度较快，不足之处在于对字符区域长宽的计算方法比较简单，精确度不高。

2. 基于形态特征的分割方法

我国车牌图像具有以下形态特征：

- 车牌长宽比固定，字符高宽比是 2:1，字符颜色和底色对比度高；
- 由多个字符组成，字符内部笔画连续，字符之间存在间隙；
- 在倾斜、扭曲、污损的情况下，仍然保持近似长方形。

根据车牌的形态特征，可以将车牌图像划分成若干具有一致性形态特征的像素区域，实现车牌字符分割。

算法 6.2　基于形态特征的车牌字符分割

输入：车牌图像。

过程：1. 搜索车牌图像，生成列曲线图 $Hst[j]$，j 为列序号：

$$Hst[j]=\begin{cases} i, f(i,j)=1 \\ iB-iT, f(i,j)=0 \end{cases}$$

$f(i,j)$代表图像在(i,j)处的灰度值，$f(i,j)=1$ 表示有字符笔画，$f(i,j)=0$ 则没有，iB、iT 分别为车牌的上下边界；

2. 删除车牌图像部分的边框，计算阈值 $T=0.4\times(iB-iT)$；

3. 通过阈值 T 分割 $Hst[j]$ 曲线，确定每个字符的分割位置。

输出：分割后的字符图像。

---

3. 基于投影法的分割方法

在二值车牌图像的垂直投影图中，搜索最优的投影点，获取当前字符宽度，以此投影点为搜索起始点，向左右两边搜索，结合垂直投影极小值点和宽度信息，实现字符分割。

基于二值化车牌图像 $D(i,j), i\in[y0,y1], j\in[x0,x1]$，该算法描述如下。

步骤 01 计算二值车牌图像的垂直投影。

$$CH(1,j)=\sum_{i=y0}^{y1} D(i,j), j\in[x0,x1]$$

步骤 02 搜索最优投影点，获取字符宽度。

在垂直投影图中，按照 j 自 $x0$ 至 $x1$ 搜索，当满足下列最优投影点条件时终止搜索：

$$\begin{cases} CH(1,j) < th_L, j \in [x_t-5, x_t-1] \\ CH(1,j) > th_H, j \in [x_t, x_t+20] \end{cases}$$

其中，th_L 为车牌投影波谷阈值，th_H 为车牌投影波峰阈值，x_t 为最优投影点。以 x_t 为起始点，分别向左右两边搜索，当搜索到第一个波峰之后的第一个波谷 x_t1 时，当前字符宽度为：

$$C_w = x_t1 - x_t$$

如果搜索完，没有找到 x_t，则修改 th_L、th_H，重复步骤 2。

步骤 03　结合字符宽度搜索其他分割点

以 x_t 为起始点，分别向左右两边搜索其余波谷，当相邻波谷点之间距离与 C_w 相差较小时，则认为当前波谷为分割点，在车牌区域内实现所有字符的分割。

基于投影法的分割方法的字符分割效果如图 6.11 所示，其中黑色竖线处为搜索到的最优投影点。

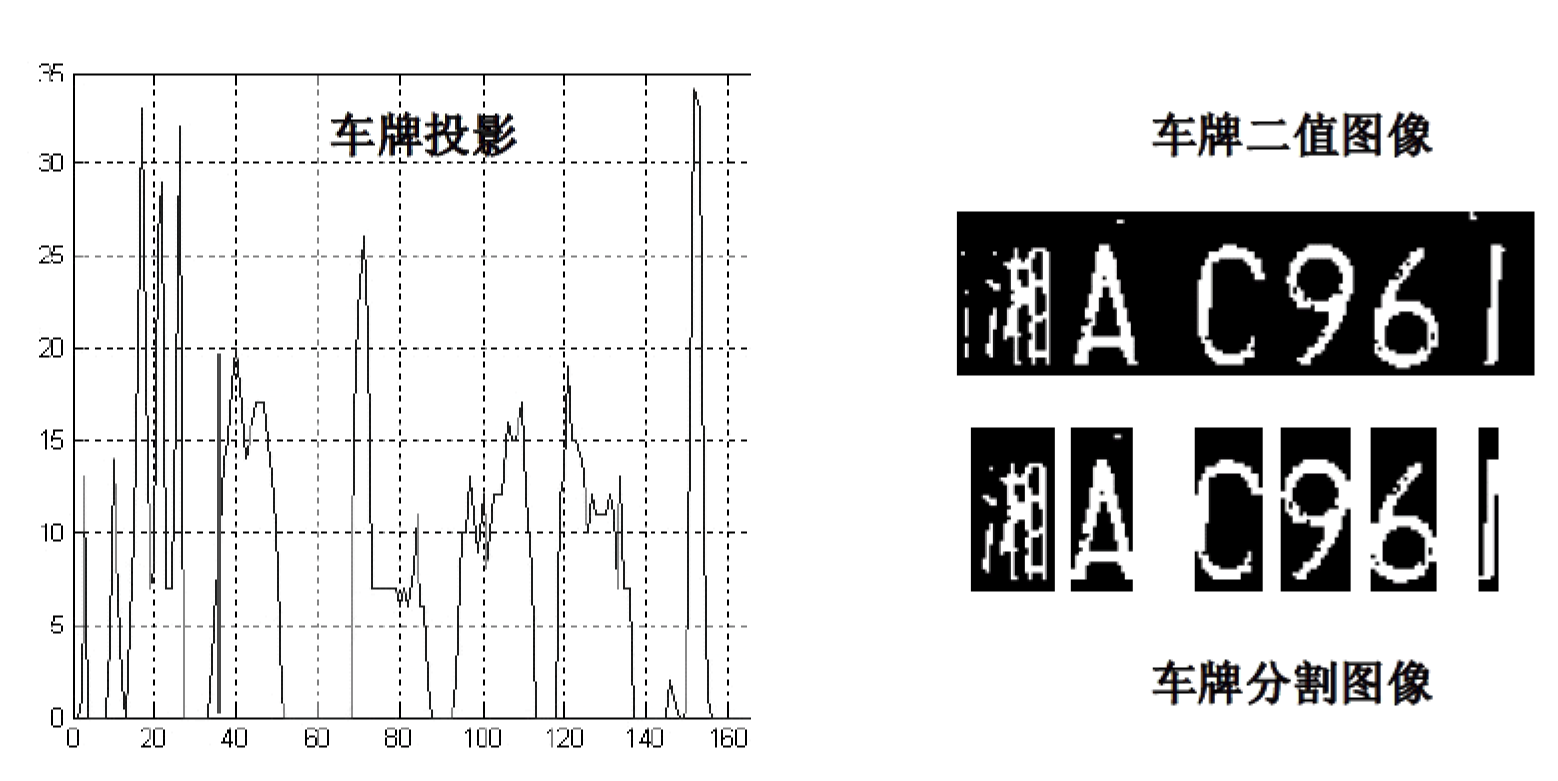


图 6.11　基于最优投影点的字符分割

6.2.4　索车牌字符识别

车牌字符识别属于模式识别和人工智能，经过分析和判断将当前字符归类为预先已知的标准字符。与其他的字符识别系统相比，我国车牌字符识别有其自身的特点，主要体现在以下方面：

- 车牌字符包括汉字、字母和数字，总计 90 个左右的字符；
- 字型统一，大小一致，相对于普通汉字识别难度较低；
- 车牌搜索要求实时性，识别算法必须保证高速度；
- 要求有较高的识别率，最低限度减小误识率。

车牌字符识别的关键，就是基于车牌搜索的具体特点，选择适用于上述特点的分类方法，常用方法有模板匹配、结构特征和统计特征等方法。

1. 模板匹配识别方法

若用 0 表示背景，1 表示字符，对汽车牌照涉及的每个字符均建立标准的模板 T_i。令待识别图像为 X，大小均为 $M\times N$，将 X 与每个标准字符模板进行匹配，分别求出它们的相似度 S_i：

$$S_i=\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}X\times T_i}{\sum_{m=1}^{M}\sum_{n=1}^{N}T_i}$$

其中，T_i和 X 均为像素的二值点阵，$X\times T_i$表示矩阵和矩阵的点乘，即矩阵中对应位置的元素相乘。上式表示标准模板和待识别字符图像对应点均为“1”像素的数目与标准模板上“1”像素的数目之比。

基于上式计算待识别字符与所有模板字符的相似度，将最高值作为其识别结果。

模板匹配方法实现简单，识别速度快，受噪声影响小；但是较难准确地提取特征，不能有效地保证识别率。

2. 结构特征识别方法

中国大陆汽车牌照中使用的字符包括 59 个汉字、25 个英文字母（I 除外）和 10 个阿拉伯数字 3 种类型，共 94 个字符，都是印刷体，结构固定、笔画规范。全部字母和数字的笔画共有两大类：直笔画和弧笔画。直笔画可分为横笔画、竖笔画、左斜笔画和右斜笔画。弧笔画是一条曲线段，可分为两类：开弧笔画和闭弧笔画。所谓开弧笔画指该弧笔画没有形成封闭环，如字母 C；而闭弧笔画则为封闭环，如数字 0。

在字符图像的结构特征中，封闭环可作为字符识别的重要依据，封闭环的搜索流程如图 6.12 所示。

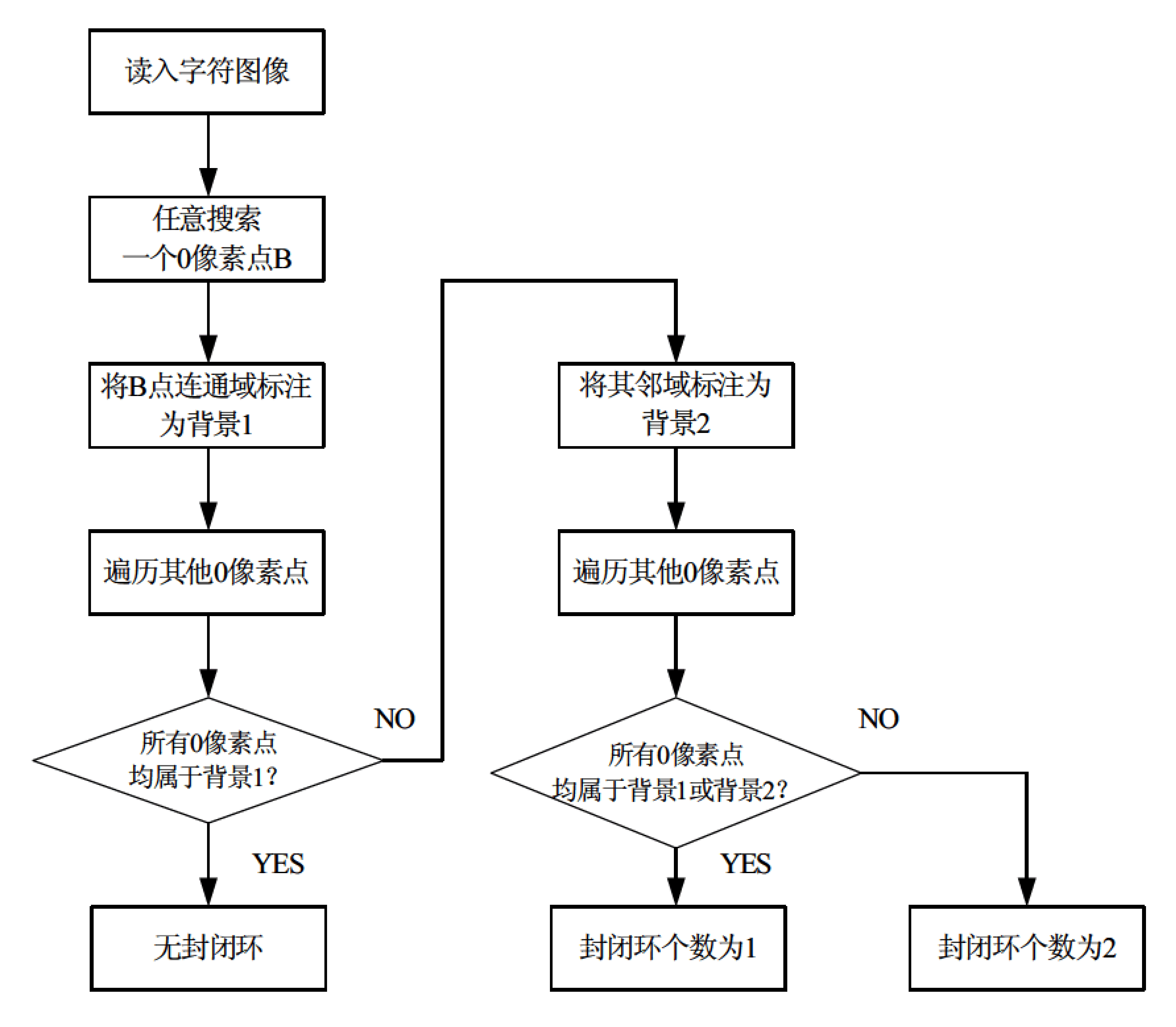


图 6.12　封闭环搜索流程

在获取到字符图像中的封闭环信息之后，可以采用知识树方法，结合封闭环以及字符笔画特征，对字母和数字逐级分类识别。

基于结构特征的字符识别方法的核心是通过判定树对字符群体层层分类，从树干开始逐步缩小识别范围，直到最后只有一类字符，即识别成功。

3. 统计特征识别方法

统计方法由于具有良好的鲁棒性和抗干扰性等优点，得到了深入研究和广泛应用。其中人工神经网络和支持向量机（SVM）在车牌字符识别研究领域取得了较好效果，能有效地提高识别率。

（1）神经网络方法

如图 6.13 所示，首先提取标准字符的特征，利用其特征训练预先设置的神经网络；然后提取待识别字符的特征；最后将特征输入人工神经网络，输出即为识别结果。

在神经网络识别方法中，神经网络的层数和各层神经元的个数，直接影响处理速度和识别正确率，层数和各层神经元的个数越多，识别正确率越高，但是这制约了识别速

度。影响成功率的因素还包括训练样本的数量以及训练次数，必须具备一定数量的训练样本和次数，才能保证识别正确率。在实际应用中应根据处理结果适当调整各种参数，兼顾识别速度与质量。

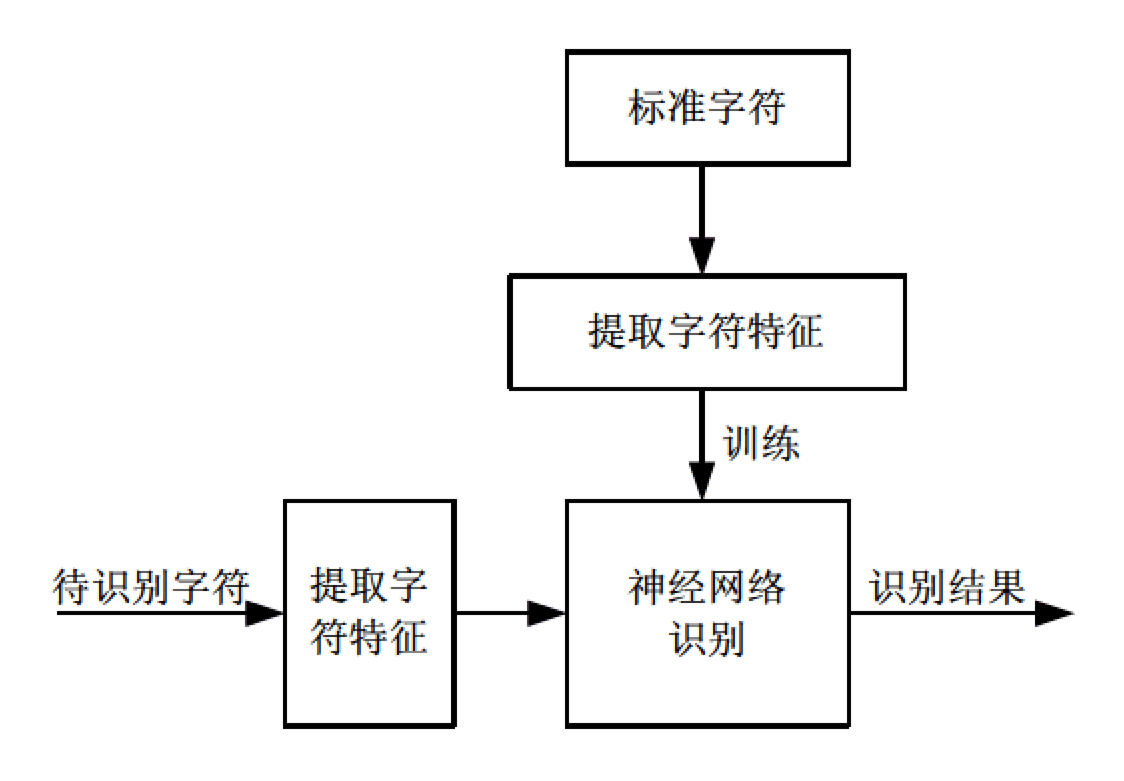


图 6.13　基于神经网络的车牌字符识别

（2）SVM 方法

支持矢量机（Support Vector Machine，SVM）利用 Vapnik 等提出的统计学习理论，能较好地弥补小样本学习和神经网络方法的不足。

SVM 针对二分类问题，车牌字符识别需要解决多分类问题，主要有 3 种方法实现 SVM 的多分类：逐一鉴别法、一一区分法和 M-ray 法。其中逐一鉴别法构造子分类器相对简单，且计算量适中，可采用该方法进行车牌搜索。

我国大陆车牌字符通常由汉字、英文字母和数字组成，如果将所有字符混在一起分类，将会降低识别率，并增加训练难度和时间。我国大陆常用车牌字符满足如下要求：第 1 个字符为汉字，第 2 个字符为大写英文字母，第 3~7 个字符为大写英文字母或数字。因此可将子分类器分为两组：汉字组分类器组、英文字母及数字组分类器。对每个分类器，首先应建立具有比较满杯的字符样本库，然后对样本数据进行训练，得到各类字符对应的判别函数。根据序号，将分割后的单个字符送到相应的分类器，各个判别函数对其进行分类，，最终输出识别分类结果。

与神经网络方法相比，SVM 不仅所需的样本少，而且泛化能力好，容易控制。

6.3　车标搜索子系统

套牌车严重影响车牌的唯一性，扰乱社会秩序，为社会安全埋下了隐患。如图 6.14 所示，车标也是车辆的显著标识，车标种类繁多，大小各异，位置不确定。大部分车标位于散热器的中心位置，少部分车标位于散热器的顶端。



图 6.14　车标示例

车标搜索与车牌搜索都是智能交通系统的重要组成部分，如图 6.15 所示，车标搜索利用图像采集设备采集车辆正面图像，由计算机对车标进行定位和识别。

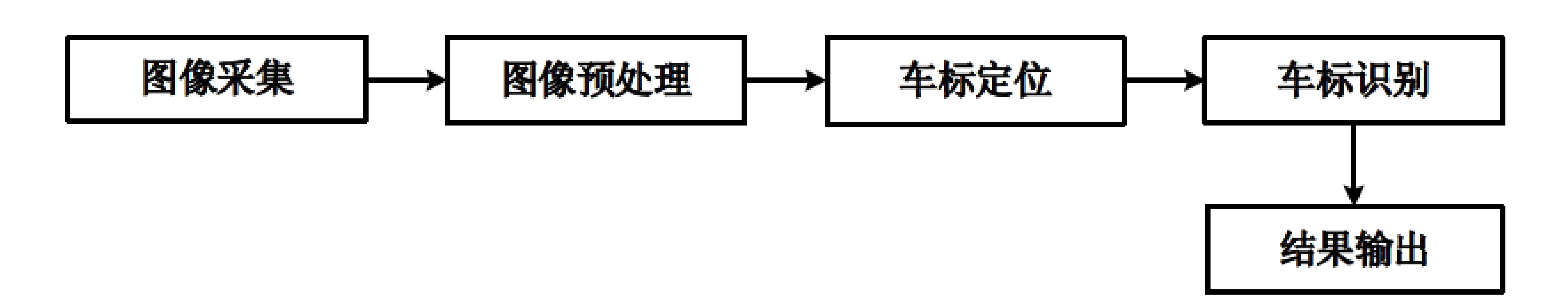


图 6.15　车标搜索流程

6.3.1　车标定位

车标定位是车标搜索的基础，如图 6.16 所示，首先借助车牌定位信息进行车标粗定位；然后对粗定位区域进行边缘检测，得到车标的边缘轮廓图；接着对边缘轮廓图进行背景判断分析，去除背景干扰；最后采用数学形态学滤波车标图像，即二值闭运算，实现车标的精确定位。

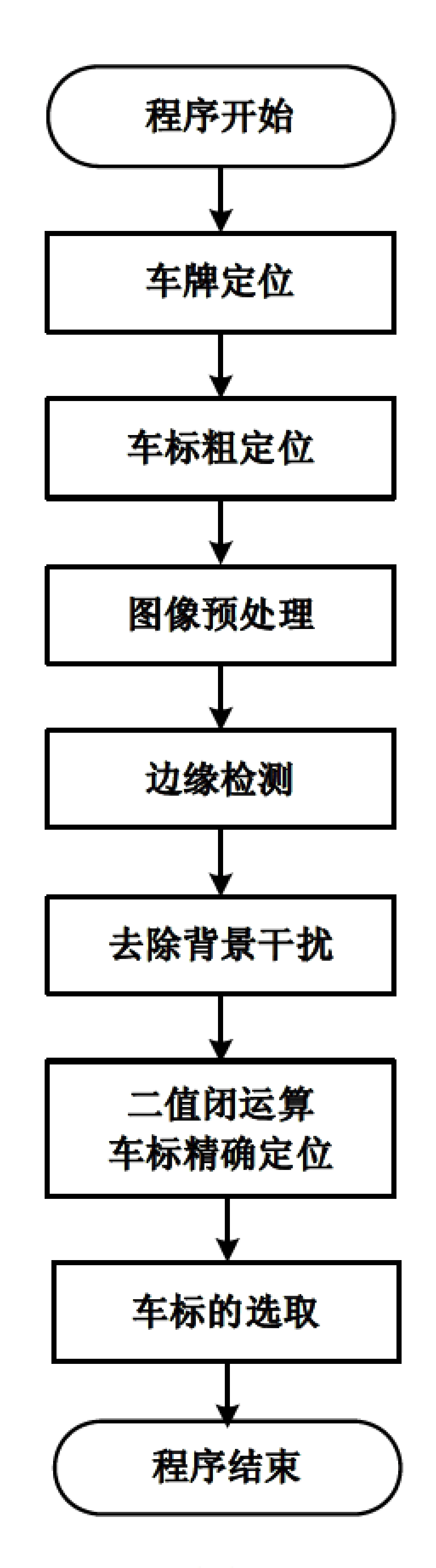


图 6.16 车标定位流程

1. 车标区域粗定位

车标一般位于中轴线之上，根据经验可以粗定位车标区域，缩小车标范围。车标粗定位块的大小为：宽度从车脸宽度的 1/3 开始，到 2/3 结束；高度通过投影确定。

如图 6.17 所示，a 为车脸图像；b 为车脸图像的中间 1/3，用作投影块；c 为对应的 Sobel 边缘图；d 为其二值图，做投影分析；e 为车标粗定位图。

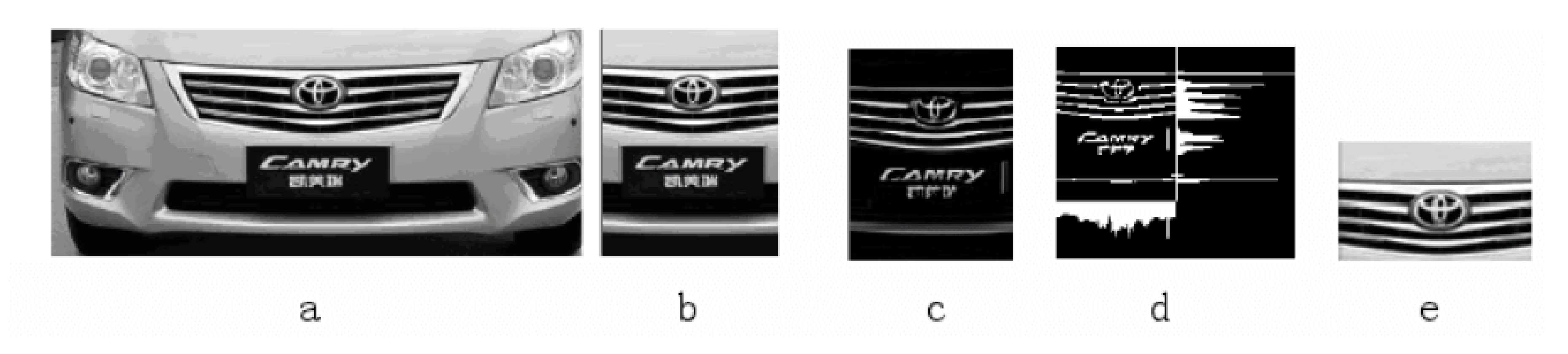


图 6.17 车标粗定位

2. 车标准确定位

车标背景大部分都是水平或垂直边缘性很强的散热片、光滑的车身表面，大多数车标在这两个方面都具有很强的边缘特性，可以在水平和垂直方向检测边缘，抑制背景，获取车标的边缘轮廓特征。sobel 边缘检测算子有垂直和水平两个方向的模板，可以用于检测图像中垂直和水平方向的边缘，精度较高，容易实现，能进一步抑制噪声的干扰。

如图 6.18 所示，以散热器为横向纹理的大众和散热器为竖直纹理的别克车标为例，分别采用 Sobel 算子的垂直模板、水平模板检测边缘。



图 6.18　sobel 边缘检测

在边缘检测之后，散热器的纹理基本滤除，但是还有少许干扰。如图 6.19 所示，对于 sobel 边缘检测后的车标图像，进行闭运算操作，采用矩形结构元素进行膨胀和腐蚀。



图 6.19　闭运算效果

在删除面积较小的孤立点之后，效果如图 6.20 所示。



图 6.20　过滤孤立点

如图 6.21 所示为车标粗定位和准确定位效果。

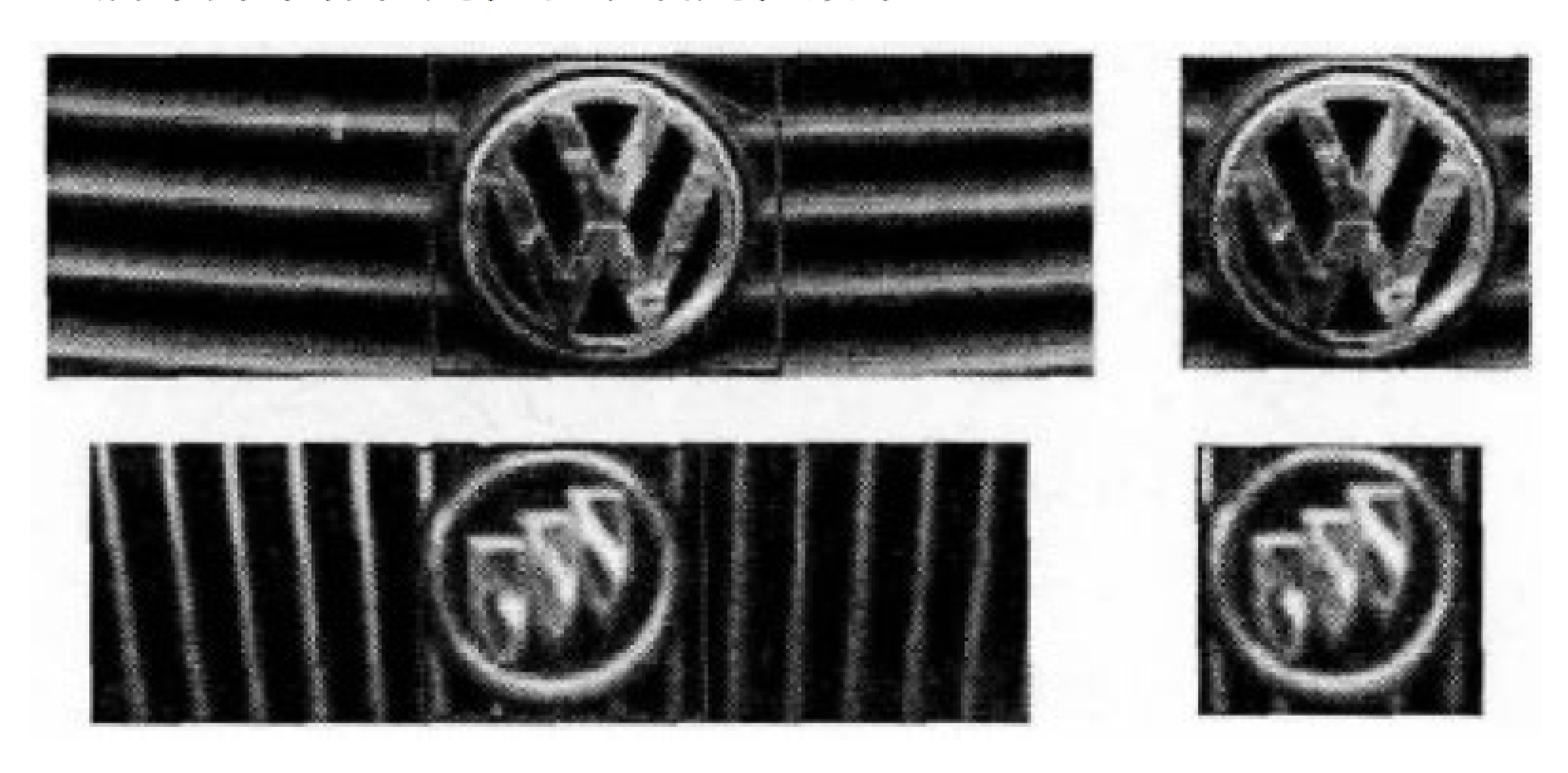


图 6.21　车标定位效果

6.3.2　车标搜索

车标搜索的关键在于如何提取出合理特征以及选择何种分类方法，车标搜索存在许多问题。

- 车辆类别增长速度快，模板库需要不断变化;
- 近似车标越来越多，难以自动区分;
- 车标区域的分辨率偏低，特征提取困难;
- 光照变化大、角度易偏差、背景复杂。

1. 基于边缘方向直方图的车标搜索方法

灰度直方图主要体现图像的灰度分布情况，受光照的影响较大；边缘方向直方图描述的是图像边缘的统计特征，可以更本质地提取目标的形状和边缘特征，受光照的影响小。

算法 6.3　车标边缘方向直方图模型

输入：车标灰度图像。

过程：1. 在灰度目标图像 $f(x,y)$ 中采用边缘算子，得到各个像素点在 X 和 Y 方向上的变化量 dx 和 dy；

2. 计算各个像素点的边缘方向角的弧值θ:

$$\theta(x,y)=\arg tg(dx/dy)$$

3. 将边缘方向角的弧值θ量化为从 0 到 T-l;

4. 将量化后的边缘方向角弧值θ进行直方图统计，得到 h(i)(i=0，1,⋯,T-1)。

输出：车标边缘方向直方图。

图 6.22 是几类车标样本的边缘方向直方图。

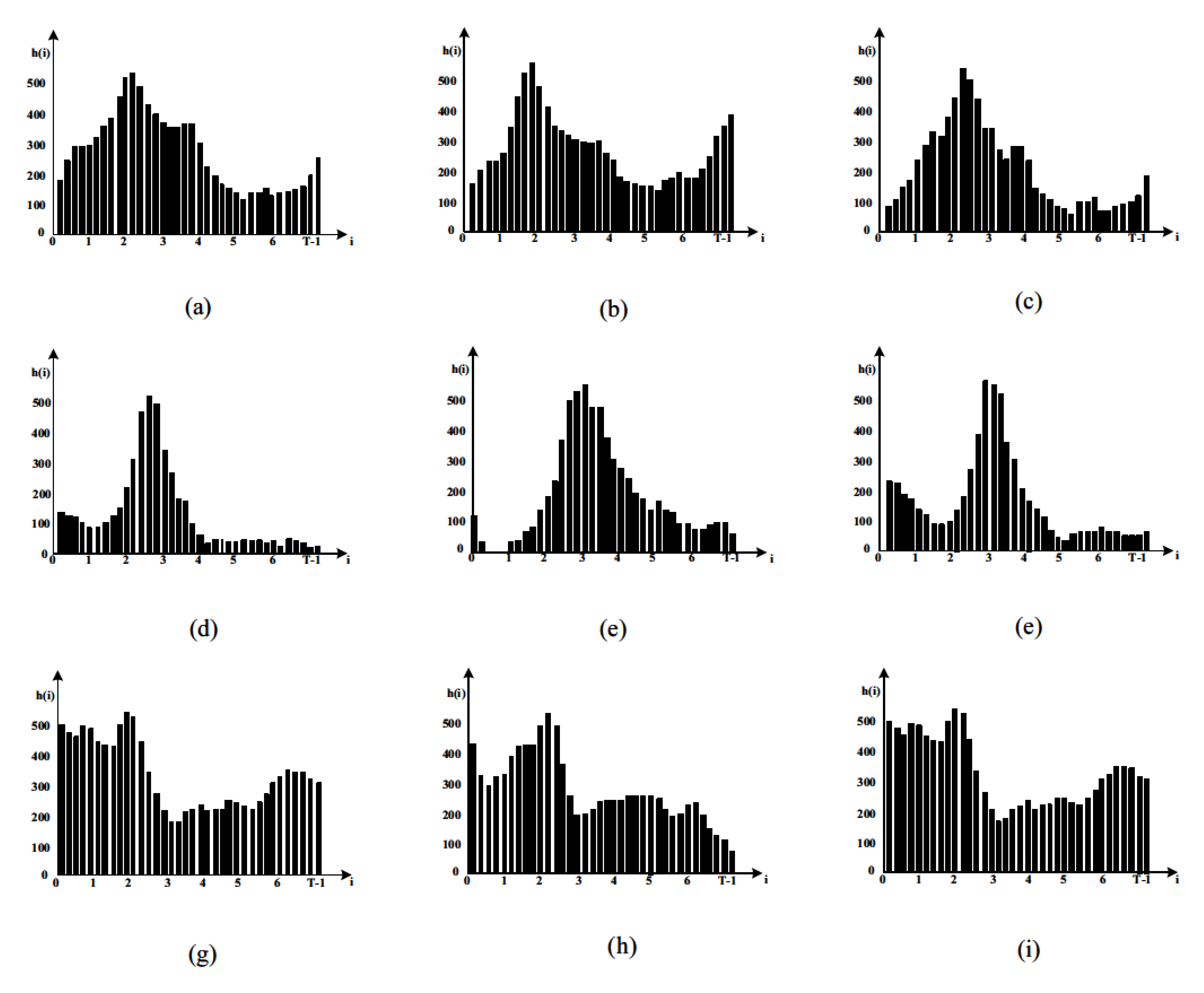


图 6.22　车标图像的边缘方向直方图

其中(a)、(b)、(c)是大众车标的边缘方向直方图，(d)、(e)、(f)是欧宝车标的边缘方向直方图，(g)、(h)、(i)是奥迪车标的边缘方向直方图。

经实验及观察发现，属于同一品牌车辆的车标，其边缘方向直方图有较强的相似性，不同品牌车标的边缘方向直方图有较大的差异性。因此可选取边缘方向直方图作为车标分类的特征。

提取待识别车标区域图像的边缘方向直方图 $h(i)_k(k=1,2\cdots,l)$，再将 $h(i)_k$ 分别与各类别标准模板的边缘方向直方图 $H(i)_k$ 进行相似性比较，采用欧氏距离衡量相似性：

$$E_k = \sum_{i=0}^{l=T-1} \left| H(i)_k - h(i)_k \right| \qquad k=1,2,\cdots,l$$

若 E_k 越小，则目标与该类别模板就越相似。

基于边缘方向直方图的特征匹配法较为简单，计算速度快，能较好反映目标图像的边缘和形状特征，各类车标特征的分离性能较好。

2. 基于 SIFT 的车标搜索方法

1999 年 David G.Lowe 提出 SIFT，即尺度不变特征变换，在总结不变特征检测方法的基础上，提出基于尺度空间的特征匹配算法，能提取稳定特征，具有平移、旋转、仿射变换、视角变换、光照变换的不变性。可应用于目标识别、图像检索、目标跟踪等领域，准确度较高。

算法 6.4　基于 SIFT 的车标搜索方法输入：车标灰度图像。

过程：1. 提取一定数量的特征点，并保存其对应的特征描述子；
2. 按行优先顺序对截取的特征点进行排序；
3. 将排序后的关键点对应的特征值连在一起，组成特征向量，其维数是 N=n×128，其中 n 为 SIFT 关键点的个数；
4. 将 SIFT 特征描述子作为车标搜索特征，输入分类器，实现搜索。

输出：车标搜索结果。

---

3. 基于 SURF 的车标搜索方法

SURF（Speed Up Robust Feature）是一种提取局部特征的算法，在尺度空间寻找极值点，使之具有尺度、旋转不变性，并对视角变化、仿射变换和噪声保持一定程度的稳定性，独特性好，信息量丰富，比 SIFT 计算量小、速度快。

SURF 实现过程主要包括关键点提取和特征描述两部分。首先基于尺度空间理论，

Bay 等人利用 Hessian 矩阵提取关键点；然后采用方框滤波（box filters）近似代替二阶高斯滤波，并用积分图像加速卷积，提高计算速度；接着检测尺度空间的极值点，精确定位极值点；最后为这些关键点指定方向。

SURF 关键点描述如图 6.23 所示。

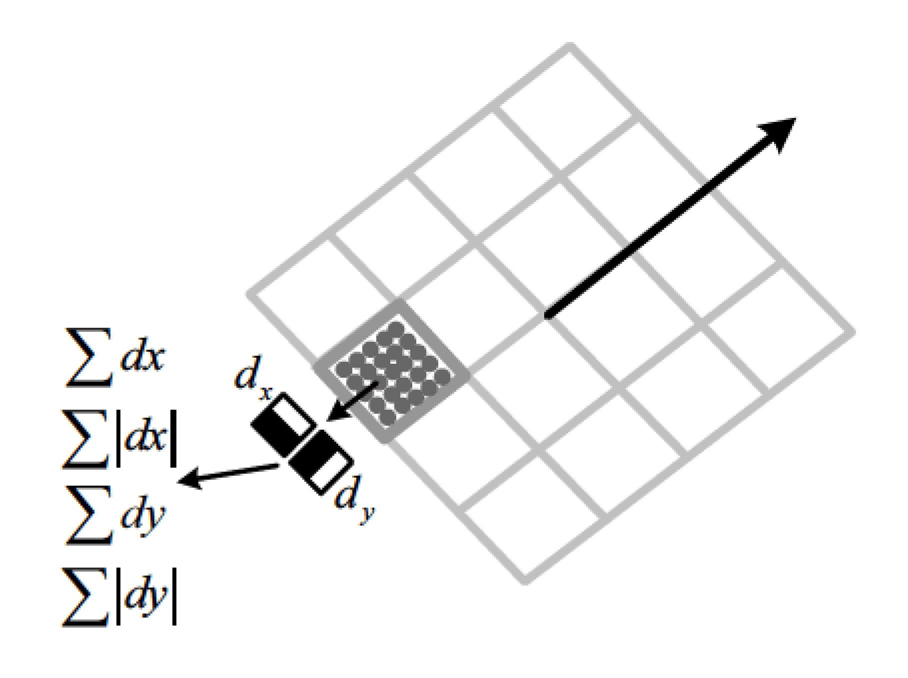


图 6.23　SURF 关键点描述

以特征点为中心，首先将坐标轴旋转到主方向；然后按照主方向选取边长为 20s（s 为当前尺度量）的正方形区域，将该窗口区域划分成 4×4 的子区域，在每一个子区域内计算 5s×5s（采样步长为 s）范围内的 Haar 小波响应。

将每个子区域的 Haar 小波响应和其绝对值相加得到$\sum dx$、$\sum dy$、$\sum |dx|$、$\sum |dy|$。因此对每个特征点，形成 4×4×4 = 64 维的描述向量，并进行向量的归一化，从而对光照具有一定的鲁棒性。

如图 6.24 所示，针对待识别车标图像，首先提取其 SURF 特征点，然后进行特征描述，最后和模板库中的特征点进行匹配。

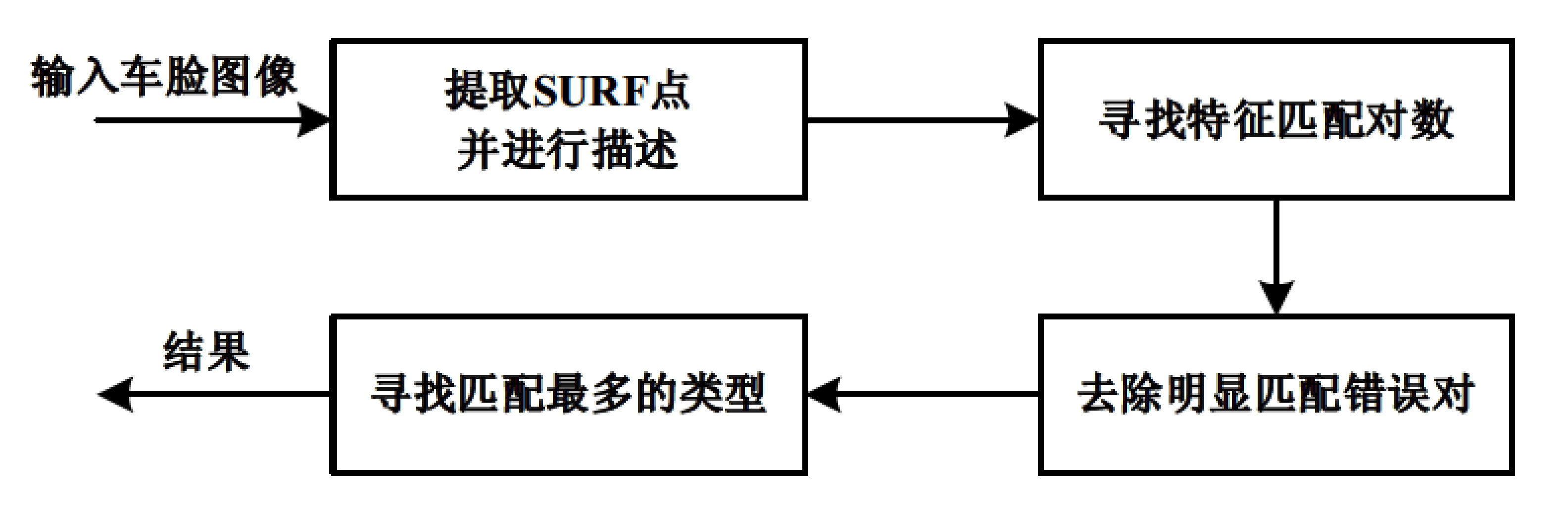


图 6.24　基于 SURF 的车标搜索

第 7 章

暴力行为检测系统

在美国“9•11”恐怖袭击之后，反恐、维稳成为世界各国非军事化行动的重要方向。视频监控系统作为主要的安全防范手段，广泛应用于银行、交通、监狱、居民小区等公共安全场所。随着高清视频监控的迅速发展，面向海量视频的实时监视、分析和报警是一个大难题，人工监视和分析无法满足高安全应用要求，智能监控需求非常强烈。暴力行为危害性大，是视频监控系统的监视重点。采用视觉计算、机器学习、人工智能等自动检测暴力行为，有助于及时发现治安和恐怖隐患，避免事态升级。

7.1　暴力行为

暴力行为是指个人或团体为达到自身目的，借助于身体、机械、武器等，发出的一种区别于正常行为的激烈而具有强制性力量的行为，以及对抗这些行为所产生的抵抗行为。暴力行为可能威胁公民的人身和财产安全（如斗殴、抢劫、追逐等治安事件），甚至威胁社会公共安全（如打、砸、抢等骚乱事件），因此，及时发现和制止暴力行为，避免暴力行为的升级，对社会和谐稳定意义重大。

如图 7.1 所示，常见的暴力行为有斗殴、打砸、抢劫和追逐等，与行走、拥抱、停留等正常行为相比，暴力行为一般具有突发性大、动作较为剧烈、不可预知等特点，具体见表 7.1。



图 7.1　暴力行为示例

表 7.1　常见暴力行为的特点

行为	特点
斗殴	人体肢体剧烈运动，或不同个体肢体之间相互交互；常伴有尖叫声、求救声
打砸	人体肢体剧烈运动，多人聚集或分散；不同人体相互运动剧烈，遮挡严重；常伴有喧哗声、物品破碎声
追逐	追逐者与被追逐目标的运动轨迹基本一致，运动速度较快，存在运动加速度
抢劫	前期表现为追逐行为，后期表现为斗殴行为；常伴有尖叫声、求救声等

7.2　暴力行为检测

暴力行为检测采用视觉机器学习方法，在计算机和嵌入式 CPU 平台上，分析监控场景的视频数据，提取暴力行为的显著性和稳健性特征，判别场景中是否存在暴力活动。

暴力行为检测涉及计算机视觉、图像处理、模式识别和人工智能等多个学科，是人体行为分析领域的重要研究方向，在公共安全监控方面具有广阔的应用前景和重大的社会效益。

7.2.1　系统框架

如图 7.2 所示，暴力行为检测系统包括用户交互模块、视频采集模块、暴力行为检测模块、视频编码与网络传输模块、数据存储与显示模块等，其核心是暴力行为检测模块。

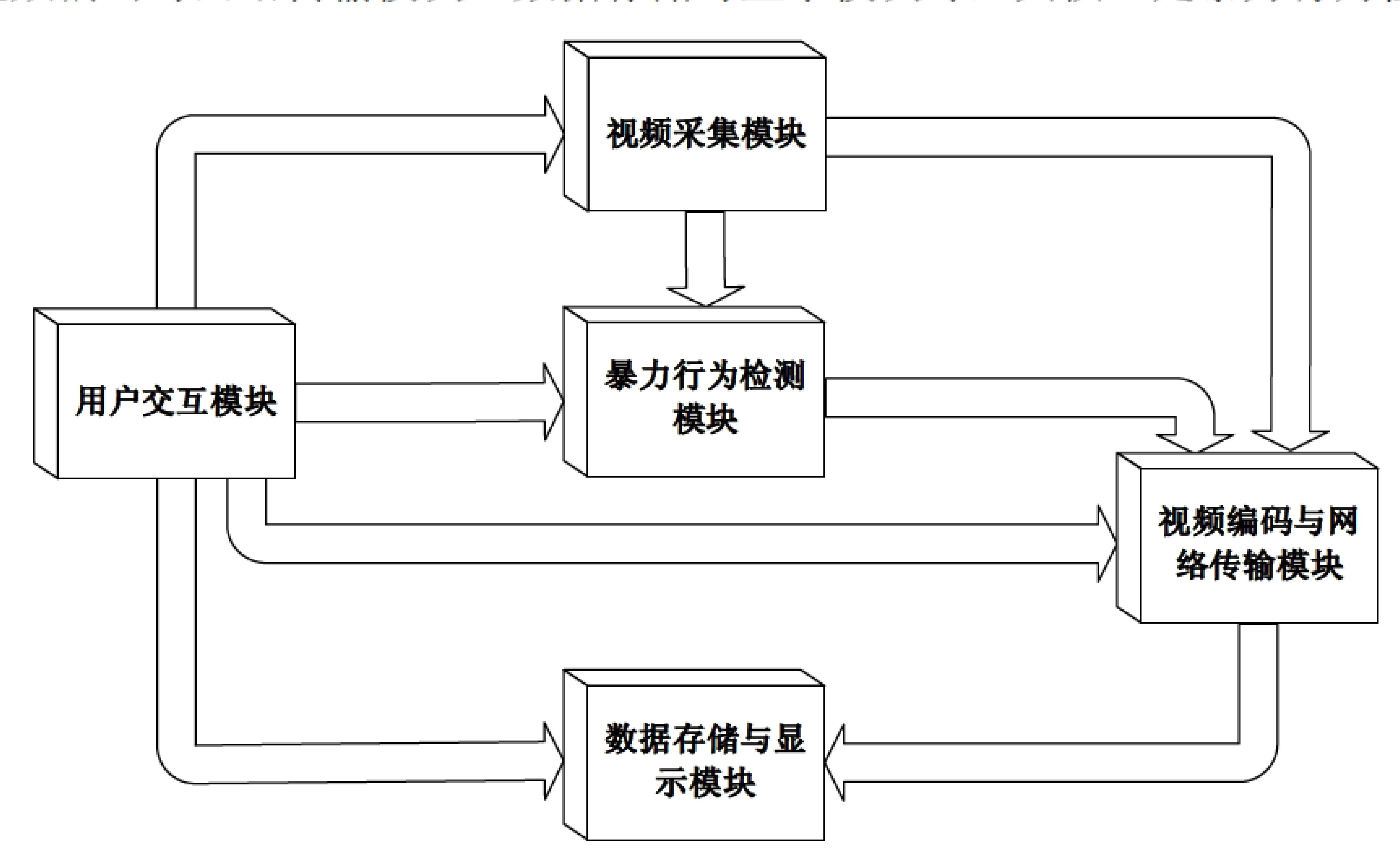


图 7.2　暴力行为检测系统框架

1. 用户交互模块

用户交互模块的主要任务是响应用户的操作指令，对系统的功能和性能进行合理配置。用户交互模块又可细分为用户登录模块和用户定制模块。

- 用户登录模块：用于验证系统管理员信息，以启动或者关闭系统，同时激活用户定制模块。
- 用户定制模块：根据用户需求，定制系统检测的功能和性能。
- 定制视频采集的性能参数，如对比度、亮度、云台转动位置等；
- 定制检测模块的性能参数，如运动检测灵敏度、判决阈值等；
- 定制视频编码与网络传输的性能参数，如帧率、码率、分辨率等；
- 定制数据存储与显示功能，如是否存储视频和报警记录、是否显示报警记录、是否进行短信报警等。

2. 视频采集模块

视频采集模块用于采集监控场景的视频。对于暴力行为检测系统而言，不仅视频质量对暴力行为检测性能有较大影响，而且摄像机的安装高度和角度也会影响暴力行为检测的性能。因此，根据场景的实际情况，需要合理地调整摄像机的高度和角度。根据应用需求的不同，可选择摄像机是否携带云台，以及选择模拟摄像机、数字摄像机或者高清摄像机。

3. 暴力行为检测模块

暴力行为检测模块采用视频分析与理解技术，检测监视视频中是否存在暴力行为。该模块是暴力行为检测系统的核心，可细分为 5 个模块。

（1）目标检测模块

目标检测模块首先通过时间差分法、背景减除法、光流法等提取前景运动目标，然后采用合理的表示方法标记和描述目标属性。其中，前景运动目标的提取是该模块的核心，算法 7.1~7.3 描述了几种常用的前景目标提取算法。

算法 7.1　时间差分法

输入：具有一定时间间隔的前一帧图像 IMG1 和当前帧图像 IMG2。

过程：1. 计算两帧图像的差值图像 IMG3：

$$f_3(x,y) = f_1(x,y) - f_2(x,y)$$

其中，(x,y)为图像中的像素点，f_1、f_2、f_3分别表示图像 IMG1、IMG2、IMG3 的颜色属性，常用亮度属性。

2. 对差值图像 IMG3 进行图像分割，得到前景与背景分离的二值图像；常用自适应阈值分割方法，如 OTSU 方法。

输出：前景与背景分离的二值图像。

算法 7.2　背景减除法

输入：背景图像 IMG1 和当前帧图像 IMG2。

过程：1. 计算两帧图像的差值图像 IMG3：

$$f_3(x,y)=f_1(x,y)-f_2(x,y)$$

其中，(x,y)为图像中的像素点，f_1、f_2、f_3分别表示图像 IMG1、IMG2、IMG3 的颜色属性，常用亮度属性。

2. 对差值图像 IMG3 进行图像分割，得到前景与背景分离的二值图像；常用自适应阈值分割方法，如 OTSU 方法。

3. 更新背景图像，常用高斯背景模型。

输出：前景与背景分离的二值图像。

算法 7.3　光流法

输入：具有一定时间间隔的前一帧图像 IMG1 和当前帧图像 IMG2。

过程：1. 计算各像素点的时间偏导 I_t：

$$I_t(x,y)=f_1(x,y)-f_2(x,y)$$

其中，(x,y)为图像中的像素点，f_1、f_2分别表示图像 IMG1、IMG2 的颜色属性，常用亮度属性。

2. 计算各像素点的空间梯度 I_g：

$$I_g(x,y)=\sqrt{I_x^2(x,y)+I_y^2(x,y)}$$

其中，

$$I_x(x,y)=f_2(x,y)-f_2(x-1,y)$$

$$I_y(x,y)=f_2(x,y)-f_2(x,y-1)$$

3. 计算各像素点的光流值 V，构建光流场：

$$V(x,y)=\frac{I_t}{I_g}$$

4. 对光流场进行分割，得到前景与背景分离的二值图像；常用自适应阈值分割方法，如 OTSU 方法。

输出：前景与背景分离的二值图像。

时间差分法对运动物体敏感，对于简单背景下的运动目标检测较为有效，对光线的变化具有较强的鲁棒性，且算法实现简单，时间和空间复杂度比较低，检测速度快，易于实时实现，在运动目标检测中应用广泛。但该方法提取的目标轮廓不完整，容易出现“孔洞”现象，检测精度较低，同时检测结果受目标运动速度的影响很大。

背景减除法检测角度较高，能够得到运动目标的完整轮廓，是目前静止摄像机视觉系统中广泛使用的方法。但该方法对光线和场景的变化非常敏感。

光流法不需要预先知道场景的任何信息，能够检测独立运动目标，即使在摄像机运动的情况下也能很好地检测出运动目标。但该方法对光线变化较为敏感，而且计算复杂耗时。

（2）目标跟踪模块

目标跟踪模块通过 Mean Shift 等方法对相邻视频帧的运动目标进行匹配，获取目标的运动轨迹、运动速度等属性。Mean Shift 的基本含义是均值偏移矢量，设 $x_1,x_2,\cdots,x_n$ 是落在以 d 维欧氏空间 R^d 点 x 为中心的单位超立方体中 S 的点集，则点 x 的均值偏移矢量的基本形式定义为：

$$M_h(x)=\frac{1}{n}\sum_{x_i\in S}(x_i-x)$$

从定义可以看出，Mean Shift 矢量 $M_h(x)$就是对落入 S 区域中的 n 个样本点相对于点 x 的偏移矢量的均值。若样本点 x_i 是从一个概率密度函数 $f(x)$中采样得到的，由于非零概率密度的梯度指向概率密度增加最大的方向，即 S 区域内的样本点更多地落在沿着概率密度梯度方向，因此，对应的 Mean Shift 矢量应该指向概率密度梯度的方向。

从表达形式可知，落入 S 区域中的 n 个样本点对均值偏移矢量 $M_h(x)$的贡献是没有差别的。为了将距离对 $M_h(x)$的影响考虑进来，在基本 Mean Shift 算法的基础上引入核函数概念，一般用 $K(x)$表示。在 Mean Shift 算法中，常用核函数有 Epanechnikov 核函数、

高斯核函数等。

Epanechnikov 核函数为：

$$K_E(x)=\begin{cases}c(1-\|x\|^2), & \|x\|<1\\ 0, & otherwise\end{cases}$$

高斯核函数为：

$$K_N(x)=c\cdot\exp(-\frac{1}{2}\|x\|^2)$$

其中 c 为 d 维单位球体的体积。

经过改进的 Mean Shift 算法实际上是一种基于核密度估计的无参数模式匹配算法，是一种计算局部最优解的实用方法。它通过迭代来搜索目标，实现对运动目标的定位，然后在视频序列上，获取目标的运动轨迹、运动速度等属性。常用的 Parzen E 核密度估计函数为：

$$P(x)=\frac{1}{n}\sum_{i=1}^{n}K(x-x_i)$$

其中

$$K(x-x_i)=c\cdot k\left(\left\|\frac{x-x_i}{h}\right\|^2\right)$$

对 p(x)取梯度，可得：

$$\nabla P(x)=\frac{1}{n}\sum_{i=1}^{n}\nabla K(x-x_i)$$

令 $g(x)=-K'(x)$，可得：

$$\nabla P(x)=\frac{c}{n}\sum_{i=1}^{n}\nabla K_i=\frac{c}{n}\left[\sum_{i=1}^{n}g_i\left(\frac{\|x-x_i\|^2}{h}\right)\right]\left[m(x)-x\right]$$

其中

$$m(x)=\frac{\sum_{i=1}^{n}x_i g_i\left(\frac{\left\|x-x_i\right\|^2}{h}\right)}{\sum_{i=1}^{n}g_i\left(\frac{\left\|x-x_i\right\|^2}{h}\right)}$$

$m(x)$表示当前窗口中所覆盖的模式点的加权中心，h 表示搜索窗口的尺寸，x 表示当前搜索窗口的中心位置，则 $M_h(x)=m(x)-x$ 表示 Mean Shift 矢量。

Mean Shift 算法的实现流程如下。

步骤01 选择搜索窗口。包括窗口的初始位置、窗口的类型（均匀、多项式、指数或高斯类型）、窗口的形状（对称的或歪斜的，旋转的、圆形的或矩形的）以及窗口的大小（超出窗口的部分被截去）等。

步骤02 计算带权重的窗口重心处位置(x,y)。

步骤03 将窗口的中心设置在步骤 2 所得重心位置处。

步骤04 返回步骤 2，直至窗口不再变化。

图 7.3 展示的是一个 Mean Shift 算法应用于二维数据局部寻优的例子。初始窗口设置为最左侧的圆形实线窗口，其形心位于圆心的靶心标志处，窗口内覆盖的所有模式点的加权重心位于形心右下角的灰色靶心标志处，于是将搜索窗口更新至以该灰色靶心标志为中心处。然后以类似方式不断更新搜索窗口，直至窗口不再移动为止。图 7.3 中的箭头表示迭代过程中产生的 Mean Shift 矢量，可以看出 Mean Shift 矢量的长度在寻优过程中不断减小，并最终收敛至极小值。

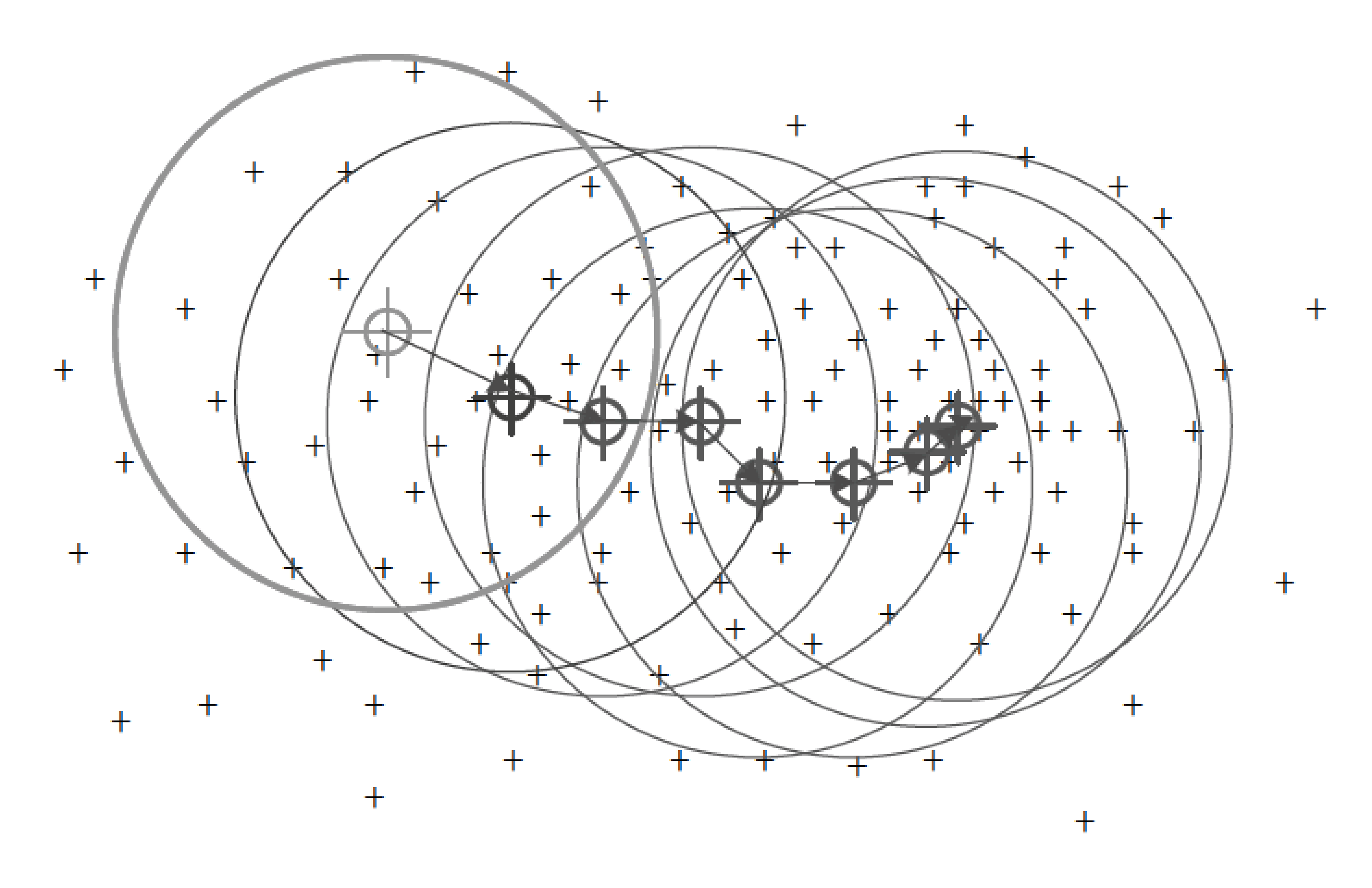


图 7.3　Mean Shift 算法局部寻优图

（3）目标分类模块

目标分类模块根据运动目标的属性特征，判断运动目标是否为人体运动目标，常用方向梯度直方图（Histogram of Oriented Gradient，HOG）特征判别人体目标。所谓 HOG 特征，是对图像矩形窗口中各个方向梯度强度的一种统计信息。Dalal 等人提出的原始 HOG 特征定义检测窗口（Detection Window）尺寸为 64×128，块（block）尺寸为 16×16，每个块包含 2×2=4 个均匀分布的 8×8 单元格（cell）。

梯度的计算方法为：

$$\begin{cases} G_x(x,y) = f(x+1,y) - f(x-1,y) \\ G_y(x,y) = f(x,y+1) - f(x,y-1) \end{cases}$$

$G_x(x,y)$ 与 $G_y(x,y)$ 分别表示样本图像在点 $f(x,y)$ 沿水平方向与垂直方向的梯度幅值。该像素点的梯度大小定义为：

$$G(x,y) = \sqrt{G_x^2(x,y) + G_y^2(x,y)}$$

梯度方向定义为：

$$\phi(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right)$$

将梯度方向空间均匀划分为对称的 9 个方向区间（bin），则每个块累积每个方向区间的像素梯度值，输出一个 4×9＝36 维的子直方图。以 8 为步长计算重叠块的方向梯度直方图，对于尺寸为 64×128 的检测窗口，除去四周边缘的 16 个像素点，则获得 7×15＝105 个 36 维的子直方图，最终组成一个 105×36＝3780 维的方向梯度直方图。

直接按定义计算 HOG 特征的复杂度比较高，在实际应用中，可以采用积分图的形式计算 HOG 特征，以大幅度降低计算量。求取过程如图 7.4 所示。

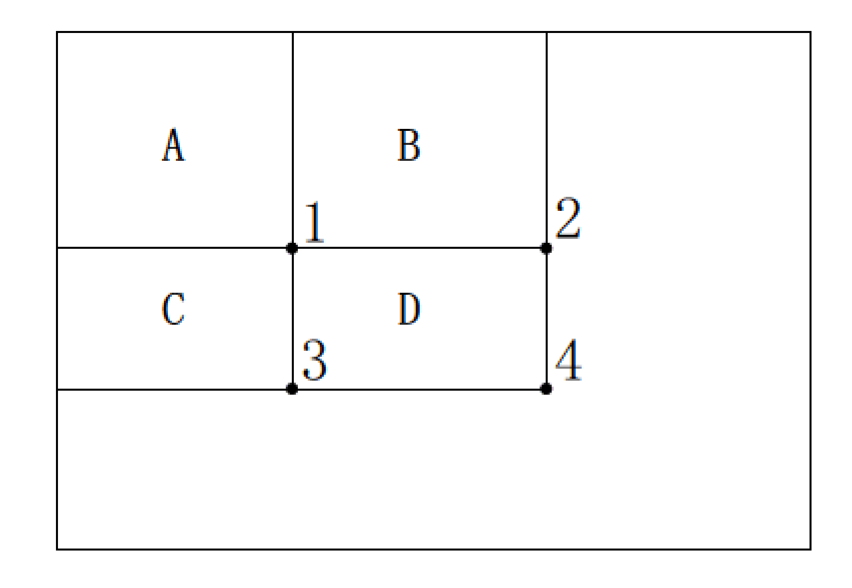


图 7.4　积分方式计算 HOG 特征示意图

定义点(x,y)的积分值为：

$$II(x',y')=\sum_{x<x',y<y'} I(x,y)$$

则图 7.4 中位置 1 处的积分值为 A，位置 2 处的积分值就是 A+B，位置 3 处积分值为 A+C，4 处的积分值为 A+B+C+D，因此可以得出矩形区域 D 的值为 D=4+1-2-3。

图 7.5 描述了 HOG 特征提取算法的实现流程，其步骤如下。

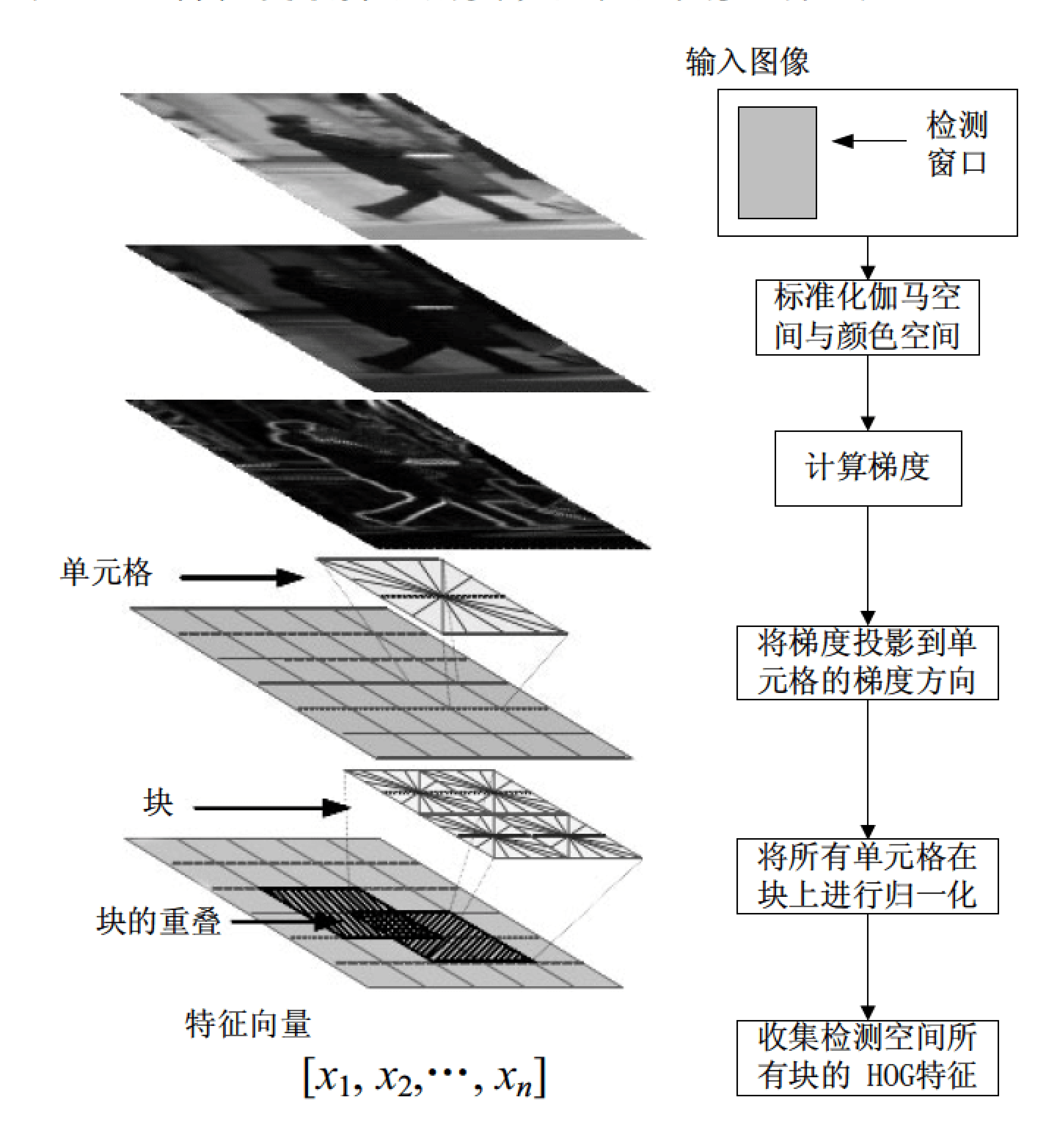


图 7.5　HOG 特征提取流程

步骤 01 为减少光照等因素的影响，将整个图像进行规范化。由于在图像的纹理强度中，局部的表层曝光的影响比重较大，因此这种压缩处理能够有效地降低图像局部阴影和光照变化的影响。

步骤 02 计算图像的一阶梯度。求导操作不仅能够捕获轮廓（contour）、剪影（silhouette）和纹理（texture）等信息，还能进一步弱化光照的影响。

步骤 03 为局部图像区域提供一个编码，同时能够保持对图像中人体对象的姿势和外观的弱敏感性。首先将图像窗口分成若干个小区域（称为单元格）；然后将每个单元格中所有像素的一维梯度直方图或者边缘方向进行累加；最后将这个基本的方向直方图映射到固定的角度上，形成方向梯度特征。

步骤04 对比度归一化。归一化能够进一步对光照、阴影和边缘进行压缩。通常，每个单元格由几个不同的块共享，但它的归一化是基于不同块的，所以计算结果也不一样。故一个单元格的特征最终会以不同的结果多次出现在特征向量中。我们将归一化块描述符称为 HOG 描述符。

步骤05 对检测窗口中所有重叠的块（overlap of blocks）进行 HOG 特征的收集，并将它们结合成最终的特征向量，供分类使用。

对于提取到的 HOG 特征，常采用 AdaBoost 算法进行训练和分类，实现流程如图 7.6 所示，具体步骤如下。

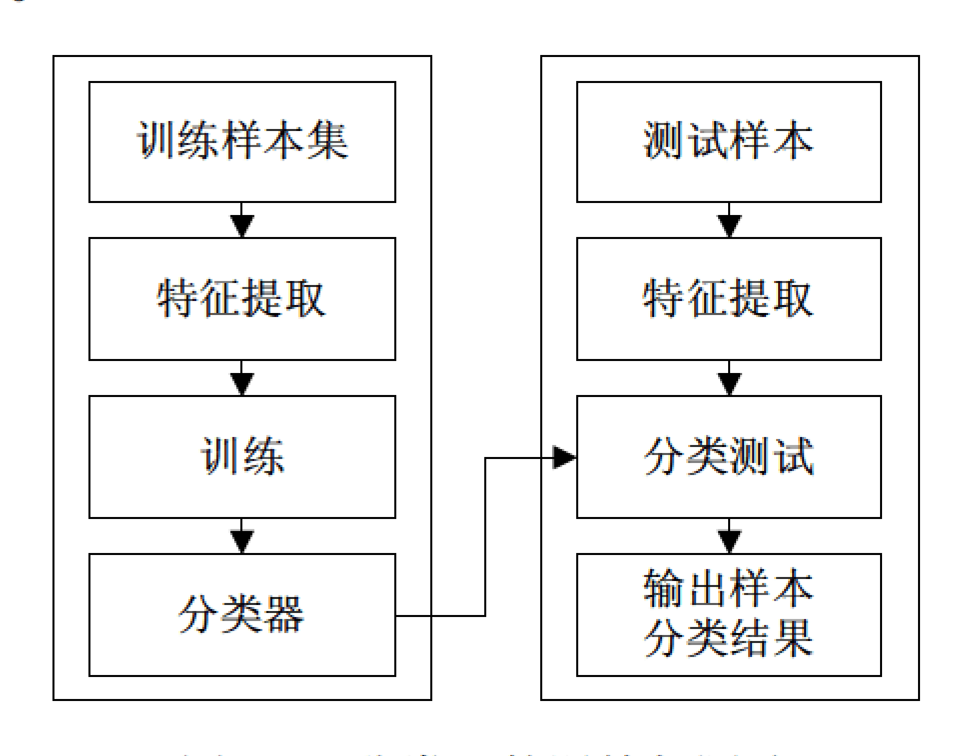


图 7.6 分类器的训练与测试

步骤01 从 INRIA 或 MIT 等数据库中选取包含人体的图像作为正样本，不包含人体的图像作为负样本，构成测试样本集，部分样本图像如图 7.7 所示，正负样本图像尺寸相同。



图 7.7 部分正样本与负样本

步骤02　提取训练样本集中每一幅样本图像的 HOG 特征。

步骤03　利用 Intel 计算机视觉开源库 OpenCV 中的 Adaboost 分类器，对步骤 2 所得数据进行训练，获得分类器。

步骤04　利用步骤 3 所得的分类器判别运动目标是否为人体目标。

（4）行为理解模块

行为理解模块依据场景中运动人体目标自身和相互之间的物理特征和运动特征等，辨别是否存在暴力行为。其中，物理特征包括不同人体目标的相对位置、人体目标各个部位的相对位置、人体目标各个部位的颜色特征等。运动特征包括人体目标各个部位的运动速度、加速度和方向等，具体的特征提取与分类方法将在后续的暴力行为检测系统介绍中详述。

（5）异常情况处理模块

当系统检测到暴力行为时，异常情况处理模块决策如何处理暴力行为，譬如进行声光报警或者短信报警等。

4. 视频编码与网络传输模块

视频编码与网络传输模块采用 H.264、AVS 等视频编码标准将监控视频进行压缩，然后融合报警信息等数据传输给远程服务器。数据的传输可以通过有线网络和无线网络进行，通过 TCP/IP 协议用 Socket 进行传输。

5. 数据存储与显示模块

数据存储与显示模块主要用于存储和显示监控场景的视频数据以及警情信息，其中涉及数据库操作、视频解码等处理。

图 7.8 所示为一个典型的暴力行为检测系统，该系统包括终端和服务器两大部分。终端主要包括暴力视频检测模块和视频编码与网络传输模块，暴力行为检测模块采用 TI DSP 实现，视频编码与网络传输模块采用 DM355 实现。监视视频信号通过视频分配器分配给 DSP 和 DM355，DSP 采用暴力行为检测算法检测监视场景中是否存在暴力行为，如果没有，则继续监视；如果有暴力行为，则通过 UART 串口将报警信号传递给 DM355。DM355 接入网线，首先实现监视视频数据的编码传输，当接收到 DSP 的报警信号后，则通过网络协议将报警信号传递给服务器，同时进行本地录像。多个终端通过网络与服务器进行通信，服务器可以实时接收各个终端的网络视频数据，实现实时监视的功能；同时，可以主动接收各个终端的报警信息，并连接数据库，自动调出终端的位置信息、负责人信息以及其他有用信息，然后通过短信平台联系负责人，及时处理终端

所在位置的暴力行为。事后，服务器还可以通过网络查询终端的录像数据，并给终端传递复位信号或参数信息。

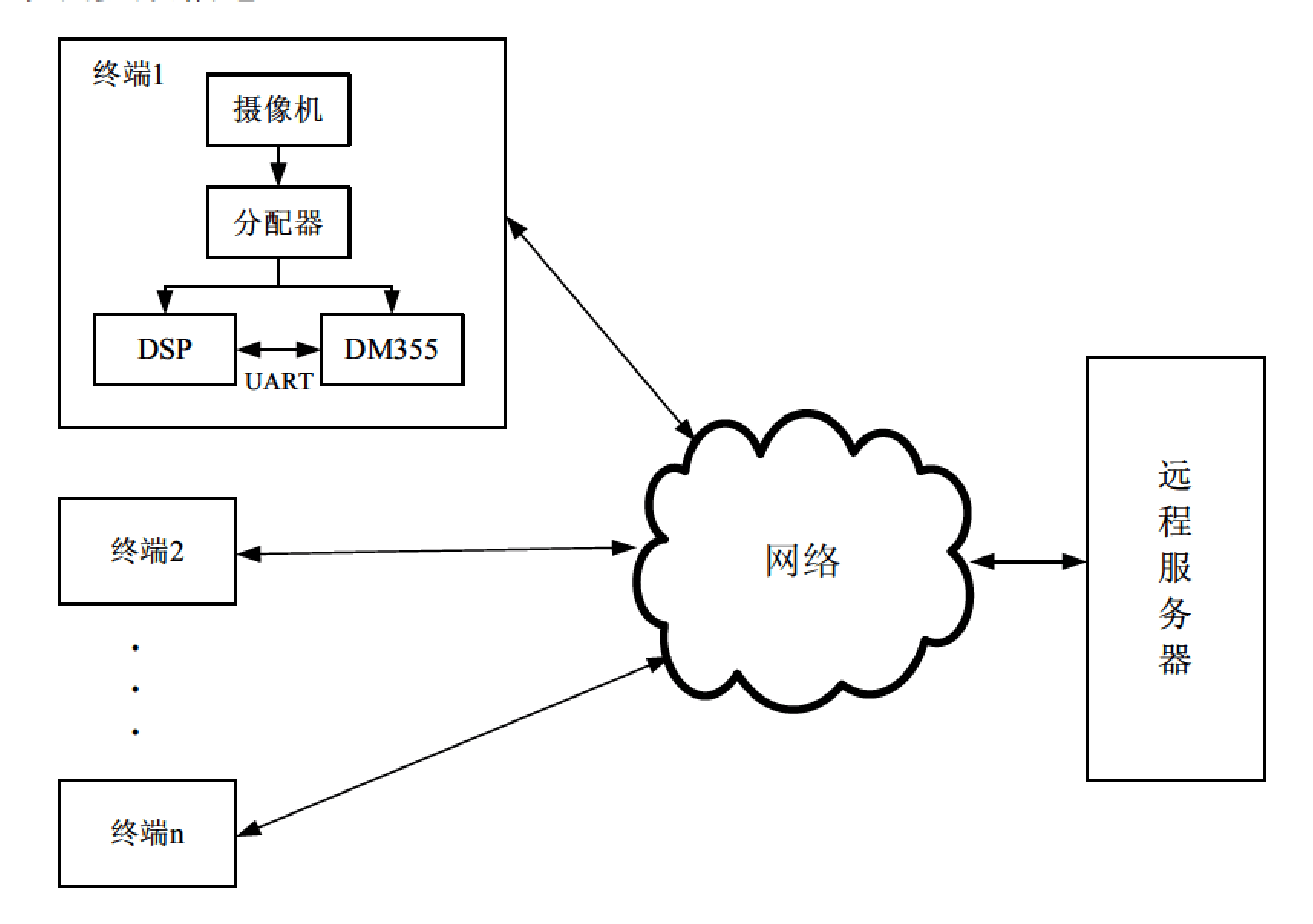


图 7.8　暴力行为检测系统组成图

7.2.2　行为数据库

视频数据库是测试和评价暴力行为检测系统性能的重要依据。国际上流行的行为识别数据库如表 7.2 所示。

表 7.2　行为识别数据库

数据库名称	创建时间	行为数量	每类行为的片段数量
KTH	2004	6	10
Weizmann	2005	9	9
IXMAS	2006	11	33
Hollywood	2008	8	30~140
UCF Sports	2009	9	14~35
Hollywood2	2009	12	61~278
UCF YouTube	2009	11	100
MSR	2009	3	14~25
Olympic	2010	16	50
UCF50	2010	50	>100
HMDB51	2011	51	>101

专门用于暴力行为检测的数据库较少，暴力行为检测系统的视频数据库一般是从这些数据库中抽取的。Bermejo 建立了一个专门用于评估打架检测的视频数据集 HockyFights，主要选取曲棍球比赛场景的运动员打架行为和正常行为，由于摄像机运动和焦距变化，给行为检测算法的测试带来很大挑战。在实际应用需求中，暴力行为特性和公共数据库中的暴力行为特性经常存在较大差异，因此，目前也有许多研究是基于自建的测试数据库进行测试和评价的。

7.2.3　评价指标

准确度、鲁棒性、速度是暴力行为识别系统的 3 个基本要求。准确度要求系统的虚警和漏警现象少，鲁棒性要求系统受噪声、光照、天气等因素的影响小，速度要求系统能满足实时监控的需求。如何选择有效的工作方案来提高系统性能、降低计算代价是异常行为识别系统值得考虑的问题。同时，如何利用来自不同用户、不同环境、不同实验条件的大量数据测试系统的实时性、鲁棒性亦相当重要。

暴力行为识别系统的定量评价指标主要有 3 个，即虚警率、漏警率和处理速度，对它们的详细介绍如下。

1. 虚警率

虚警率（FAR）是指在一定时间内，正常行为被误检为暴力行为的次数（N_1）与检测总次数（N）的比值。

$$FAR = \frac{N_1}{N} \times 100\%$$

虚警率越小，系统性能越好；反之，系统性能越差。

2. 漏警率

漏警率（FRR）是指在一定时间内，异常行为被误检为正常行为的次数（N_2）与检测总次数的比值。

$$FRR = \frac{N_2}{N} \times 100\%$$

漏警率越小，系统性能越好；反之，系统性能越差。

3. 处理速度

处理速度（FPS）是单位时间内（如 1s）系统可以处理的视频帧数（N_F）。

$$FPS = N_F$$

处理速度越快，系统性能越好；反之，系统性能越差。

暴力行为一般与行为发生场景和上下文有关，因此暴力行为识别系统的评价也与场景有关。实时监控系统对于系统的处理速度要求较高，要求暴力行为识别系统能够实时检测视频；人工监视和智能监视相结合的系统对于漏警率指标要求高，要求尽可能减少漏警，这样尽管虚警率有所增加，可以通过人工监视的方式剔除虚警，另外，由于暴力行为破坏性大，需要及时有效预警，因此暴力行为检测系统一般要求尽可能地没有漏警；自动监视系统对于虚警率指标要求较高，一般要求尽可能没有虚警，这样可以减少人力、物力的浪费。

7.3 基于对象层次的暴力行为检测系统

Ankur Datta 等人针对电影分级中的暴力行为检测问题，设计了一种基于对象层次的暴力行为检测系统。如图 7.9 所示，该系统首先求取某一对象的运动轨迹，计算具有运动方向和大小的加速度度量参数（Acceleration Measure Vector，AMV）和冲撞系数（jerk）等行为特征；然后结合周围对象的肢体运动特性，综合判决是否存在暴力行为。

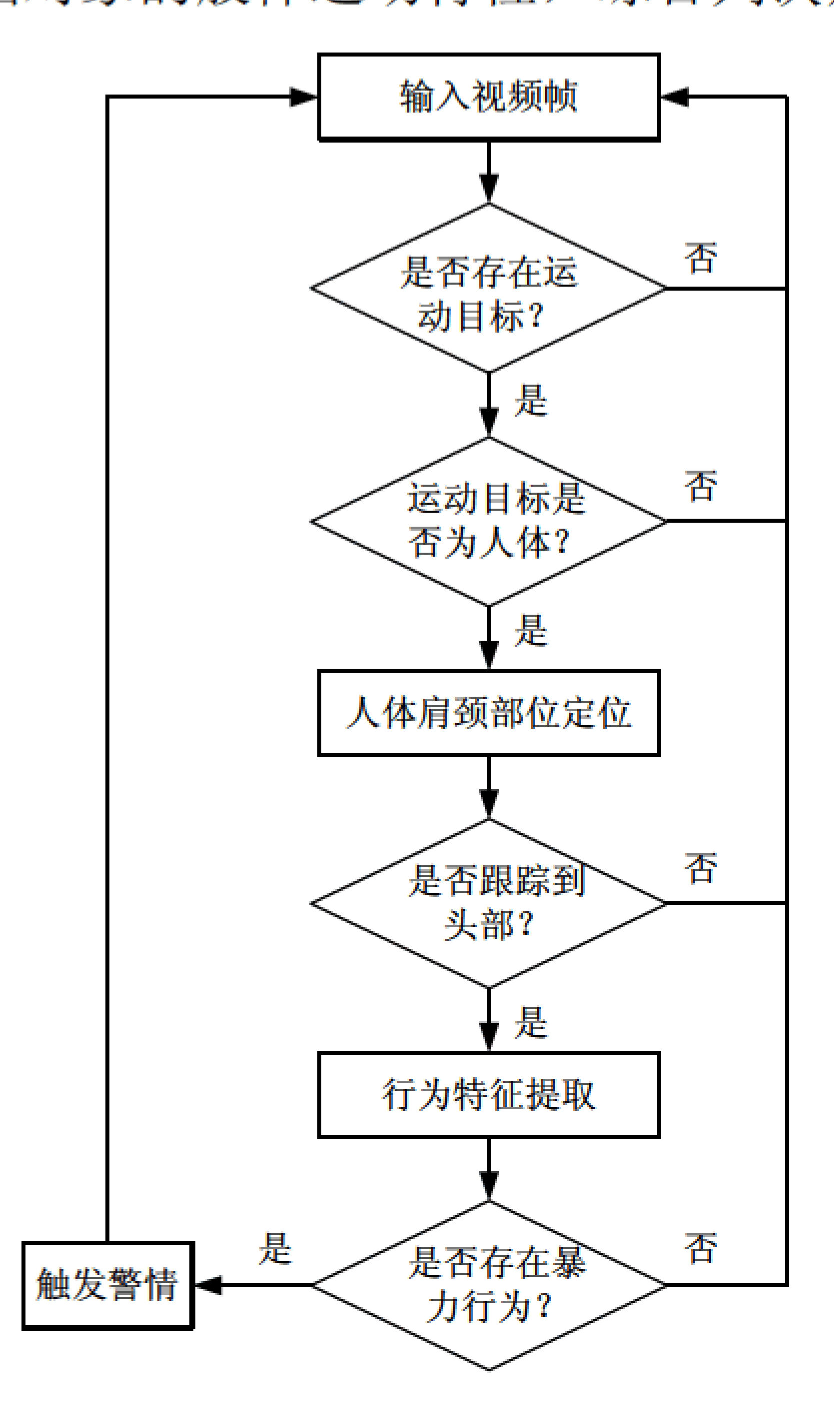


图 7.9 基于对象层次的暴力行为检测系统实现流程

1. 运动目标检测

暴力行为是由运动目标引起的，因此，首先需要检测场景中是否存在运动目标，只对存在运动目标的场景进行暴力行为检测。

常用背景减除法检测运动目标，采用单高斯模型自适应更新背景，以适应场景中的光照变化和周期运动，实现伪代码如下。

算法 7.4　单高斯模型背景减除法

输入：n 帧背景图像和当前帧图像。

过程：1. 初始化背景模型：

$$\begin{cases} \mu_i = \dfrac{1}{n}\sum_{t=1}^{n}\mu_{it} \\ \Sigma_i = \dfrac{1}{n}\sum_{t=1}^{n}E[(\mu_{it}-\mu_i)(\mu_{it}-\mu_i)^T] \end{cases}$$

其中，μ_i 为背景中任一点 i 的颜色值的期望，Σ_i 为颜色值的分布的协方差矩阵，μ_{it} 为点 i 在第 t 帧图像中的颜色值，所有背景点的 (μ_i,Σ_i) 构成初始的背景模型。

2. 目标检测：设当前帧图像上任一点 i 的颜色值为 X_i，若 $|X_i-\mu_i| < D\cdot\sigma(D\le 3)$，则认为该点为背景点，否则为目标点。式中 D 由噪声的峰度决定，一般取经验值 2.5~3。

3. 模型更新：

$$\begin{cases} \mu_{t+1} = (1-\alpha)\mu_t + \alpha X_t \\ \Sigma_{t+1} = (1-\alpha)\Sigma_t + \alpha(X_t-\mu_t)(X_t-\mu_t)^T \end{cases}$$

其中，更新率 $\alpha(0<\alpha<1)$ 是表示更新快慢的常数。

输出：前景与背景分离的二值图像。

---

2. 人体目标判定

对于检测到的运动目标，可能是人体，也可能是动物或者车辆等，而只有人体目标才能引起暴力行为，因此需要对目标属性进行判定。采用轮廓特征判定目标属性，实现伪代码如下。

算法 7.5　人体目标判断方法

输入：目标轮廓。

过程：1. 将运动目标的轮廓水平划分为 3 个相等的部分，从上到下依次记为 H1、

H2 和 H3。

2. 对于每一个部分，计算轮廓图像的垂直投影直方图。

3. 分别提取 3 个投影直方图目标数量的均值、目标数量的标准差以及轮廓边界矩形的长宽比等轮廓特征。

4. 依据正常人体目标所训练的轮廓特征，采用固定阈值法判断当前轮廓是否为人体目标的轮廓。

输出：目标是否为人体。

3. 人体颈肩部位定位

图 7.10 所示为 H1 的垂直投影直方图，依据人体颈部的先验信息，颈部的 y 轴坐标 N_r 应为直方图的谷点，也即投影的导数的极大值。依据颈部的位置，可以推算肩部的位置 S_{yy}：

$$S_y = N_r + (\frac{r}{2})$$

$$r = \sqrt{\frac{A(head)}{\pi}}$$

其中，r 表示头部的半径，$A(head)$表示头部的面积。

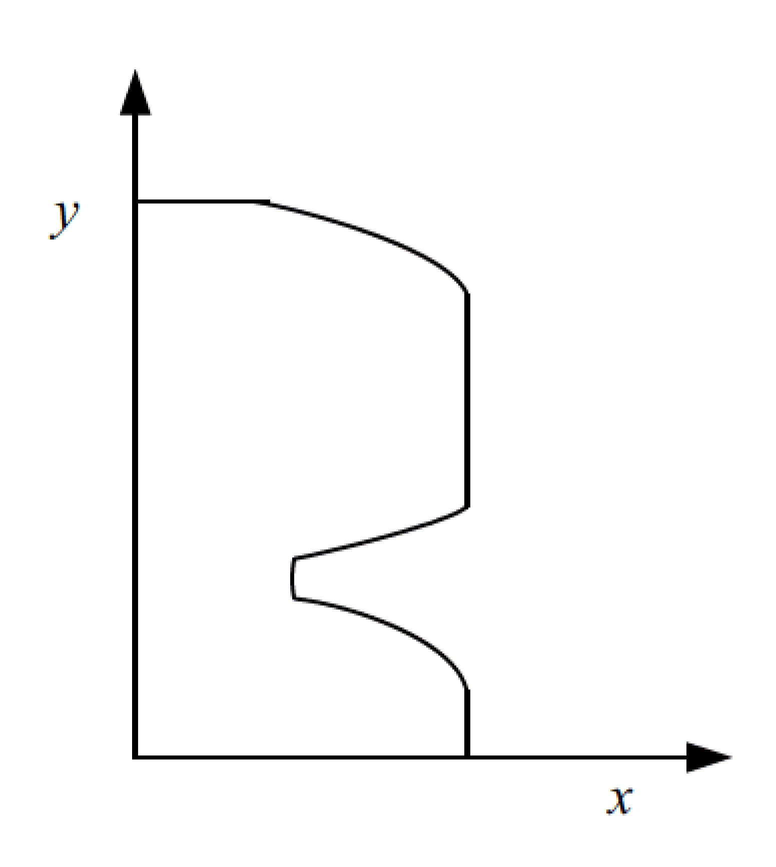


图 7.10　H1 的垂直投影直方图

4. 头部跟踪

头部跟踪的目的是获取人体目标的运动轨迹，为后续行为特征的提取服务。头部跟踪的步骤如下：

步骤01　依据上一步定位的颈部位置，采用区域生长方法向上搜索，得到头部区域。

步骤02　以头部区域的外接矩形框作为头部的初始跟踪框。

步骤03　采用颜色差分平方和（Color Sum of Squared Differences，CSSD）作为度量准则，在前一帧头部位置的附近邻域匹配当前帧中头部的位置，从而得到头部的运动轨迹。

5. 行为特征提取

依据人体头部的运动轨迹，提取具有运动方向和大小的 AMV 和 jerk 等行为特征。jerk 实际上是速度的二阶导数，可以反映暴力行为中目标运动轨迹的剧烈变化，其公式为：

$$A(t) = dV / dt$$

$$J(t) = dA / dt$$

其中，V 是速度，$A(t)$是加速度，$J(t)$即为 jerk，t 为时间。

设 $MT_i = \{\overline{P_1}, \overline{P_2}, \cdots, \overline{P_n}\}$ 表示第 i 个人的运动轨迹，其中，$\overline{P_i} = (x, y)$ 是第 i 个头部跟踪框的质心。AMV 可以定义为：

$$\delta(\theta, d) = \alpha \cdot \psi(\overline{P_{k-1}}, \overline{P_k}, \overline{P_{k+1}}) i + \beta \cdot \vartheta(\overline{P_{k-1}}, \overline{P_k}, \overline{P_{k+1}}) j$$

其中，α、β 是分别分配给加速度方向和大小的权重，d 为像素间的距离。

$$\psi(\overline{P_{k-1}}, \overline{P_k}, \overline{P_{k+1}}) = 1 - \cos\theta$$

$$\cos\theta = \frac{(\overrightarrow{P_{k-1}P_k})(\overrightarrow{P_k P_{k+1}})}{\left\|\overrightarrow{P_{k-1}P_k}\right\| \cdot \left\|\overrightarrow{P_k P_{k+1}}\right\|}$$

$$\vartheta(\overline{P_{k-1}}, \overline{P_k}, \overline{P_{k+1}}) = \left| \left(\left\|\overrightarrow{P_{k-1}P_k}\right\| - \left\|\overrightarrow{P_{k+1}P_k}\right\| \right) \right|$$

$$jerk = \sqrt{\left(\frac{\partial \psi}{\partial t}\right)^2 + \left(\frac{\partial \vartheta}{\partial t}\right)^2}$$

6. 暴力行为判定

如果第 i 帧的某个人在某方向上移动时突然改变运动方向和大小，那么这个人是遭受打击或撞击的候选人，这一现象可以采用固定阈值法用 AMV 和 jerk 来判断。如果该候选人附近有其他人，且其他人的四肢向候选人伸出，说明有人体目标正在实施暴力行为，判定该帧发生暴力行为。而四肢的方向可以采用下面的方法计算：

- 从人体的头部向肩部移动，穿过轮廓边界即可得到上臂的方向。
- 搜索 H2 部位轮廓横截面的外部边界，得到腿的方向。

当手或者腿的方向接近与地面平行或与地面成负角度时，判断该人的四肢向他人伸出。

为了减少虚警，一般采用时间滤波方法，当连续多帧图像都检测到暴力行为时，才触发警情。

7.4　基于光流变化的暴力行为检测系统

Kentaro Hayashi 等人针对电梯中的暴力行为实时检测问题，设计了一种基于光流变化的暴力行为检测系统。该系统基于光流变化提取暴力行为程度（Violent Action Degree，VAD）特征，作为暴力行为判决的依据。该系统实现简单，可以应用于实时性要求较高的电梯安全监控场所。

1. 光流变化与暴力行为的关系

光流（Optical Flow）是空间运动物体在观测成像面上的像素运动的瞬时速度，代表了局部运动的方向和模值。一般地，相对于普通行为，暴力行为发生时光流变化更大。图 7.11 为玻璃墙电梯中的两幅图像，左边一幅图像为非暴力场景，右边一幅图像为暴力场景。图 7.12 显示了图 7.11 所示场景对应的光流直方图，其中光流的计算使用典型的块匹配(Sum of Absolute Difference，SAD)方法，光流直方图划分为 8 个方向和 4 个数量级。从图 7.12 中不难发现，暴力行为发生时的光流变化比正常行为时的光流变化要大得多。图 7.13 显示了光流变化的标准偏差，其中，纵坐标代表了光流方向的标准偏差，横坐标代表了模值的标准偏差。可见，暴力行为发生时，光流方向和模值的偏差都大于正常行为。尽管不同类型的电梯的尺寸、光照、摄像机安装位置和角度不同，但暴力场景和非暴力场景在光流变化方面的属性基本不变。因此，可利用光流的变化检测暴力行为。

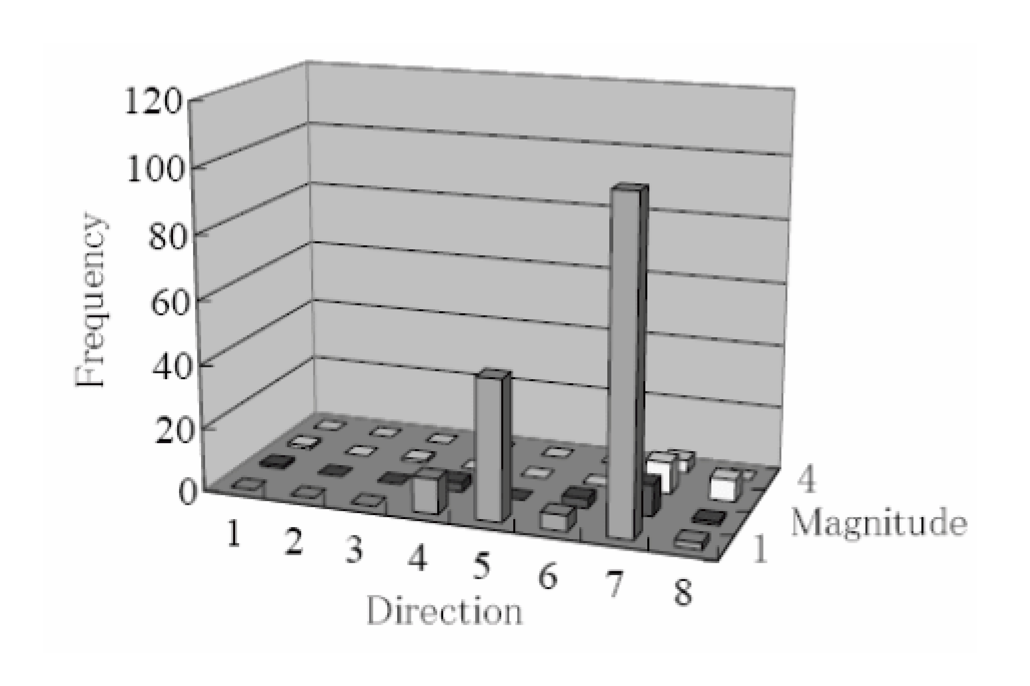


（a）非暴力场景

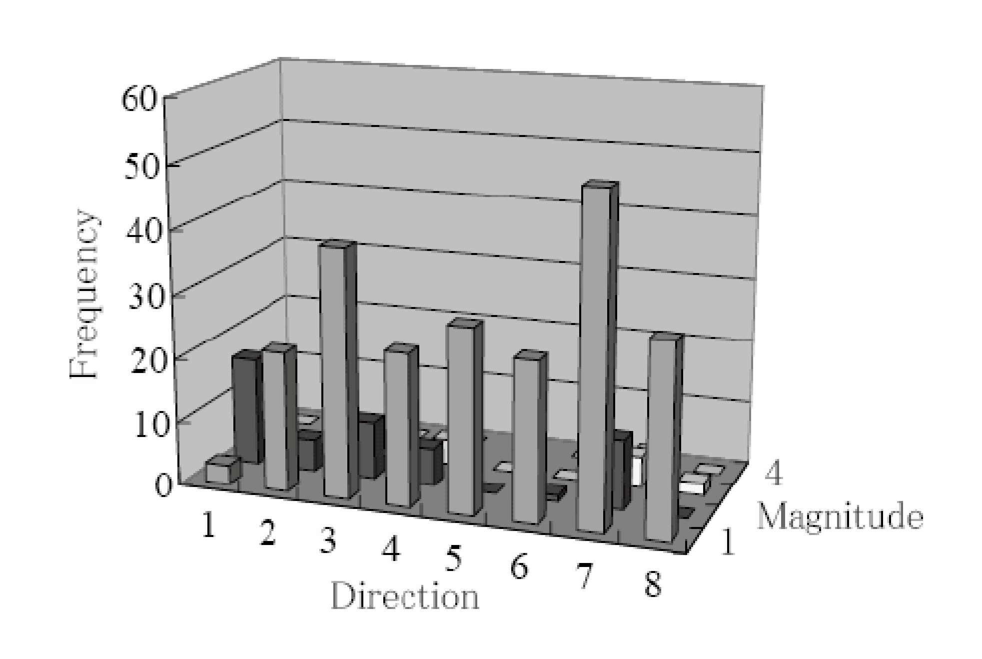


（b）暴力场景

图 7.11　电梯内的监控图像

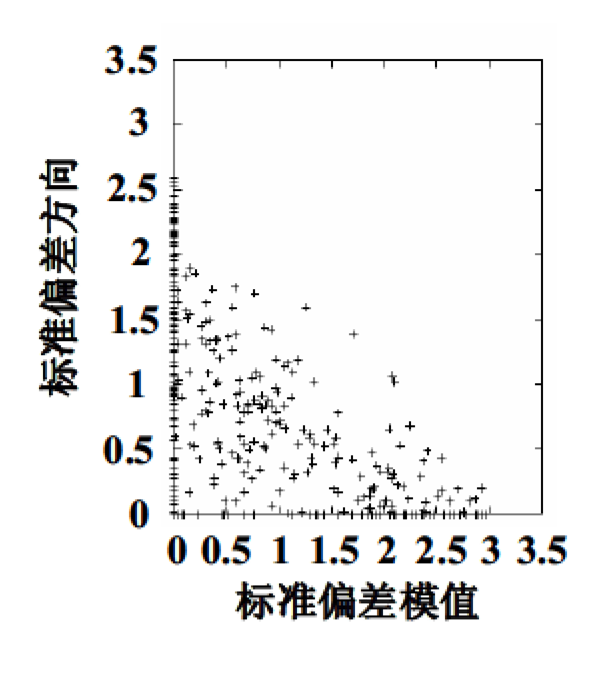


(a) 非暴力场景

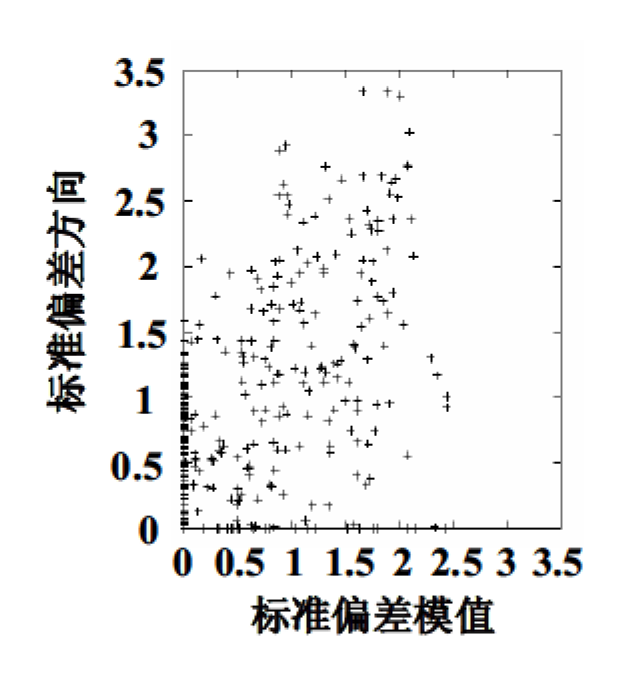


(b) 暴力场景

图 7.12　光流直方图



(a) 非暴力场景



(b) 暴力场景

图 7.13　光流变化标准偏差

2. 暴力行为检测

暴力行为的检测包括两个阶段，第一阶段是学习阶段，通过输入图像建立光流变化的模型；第二阶段是检测阶段，通过当前的光流变化和光流变化模型检测场景中是否存在暴力行为。具体实现方法描述如下。

算法 7.6　学习光流变化模型

输入：非暴力场景视频序列。

过程：1. 采用块匹配法，计算每帧图像上所有点的光流方向和模值；

2. 采用固定阈值法，剔除模值太小的光流；

3. 计算每帧图像中的 3 个光流变化特征：光流方向标准偏差 D、光流模值标准偏差 M 和光流参数的数量 Q；

4. 计算所有图像中各光流变化特征的均值和标准差：$\overline{D}$、$\overline{M}$、$\overline{Q}$、σ_D、

σ_M 和 σ_Q，此即为光流模型的内容。

输出：光流模型。

算法 7.7 暴力行为检测

输入：当前场景视频流。

过程：1. 采用块匹配法，计算当前帧图像上所有点的光流方向和模值；

2. 采用固定阈值法，剔除模值太小的光流；

3. 计算当前帧图像中的 3 个光流变化特征：光流方向标准偏差 D、光流模值标准偏差 M 和光流参数的数量 Q；

4. 通过下列光流模型计算 VAD：

$$P_d = \frac{D - \overline{D}}{\sigma_D}$$

$$P_M = \frac{M - \overline{M}}{\sigma_M}$$

$$P_q = \frac{Q - \overline{Q}}{\sigma_Q}$$

$$VAD = P_d \times P_M \times P_q$$

5. 当 VAD 超过设定的判决阈值时，判定当前帧为暴力帧；

6. 当一定时间内多帧图像为暴力帧时，判定当前场景发生暴力行为，触发警情信号。

输出：警情信号。

7.5 基于运动着色的暴力行为检测系统

在复杂监控场景中，人体目标的完整轮廓不易提取，且不同人体之间会发生遮挡现象，此时暴力行为的检测非常困难。为了解决复杂环境下的暴力行为检测问题，Alessandro Mecocci 等人设计了一种基于时空着色的暴力行为检测系统。

当发生暴力行为时，由于场景中人与人之间相互影响，会导致人体一些部位出现高速运动和局部紊乱现象。此外，因为人与人的距离更加接近，人体部位的局部变化趋向于隐蔽和不隐蔽的情况会经常发生，从而导致这些部位表面的颜色变化剧烈。因此，可

以通过分析每个场景中运动部位的时间和空间的行为着色问题，来解决暴力行为的识别问题。这里所说的运动区域，并不需要指向所涉及目标的真实部位或者轮廓，因此该系统容易处理人体遮挡问题。

该系统的实现流程如图 7.14 所示，详细描述如下。

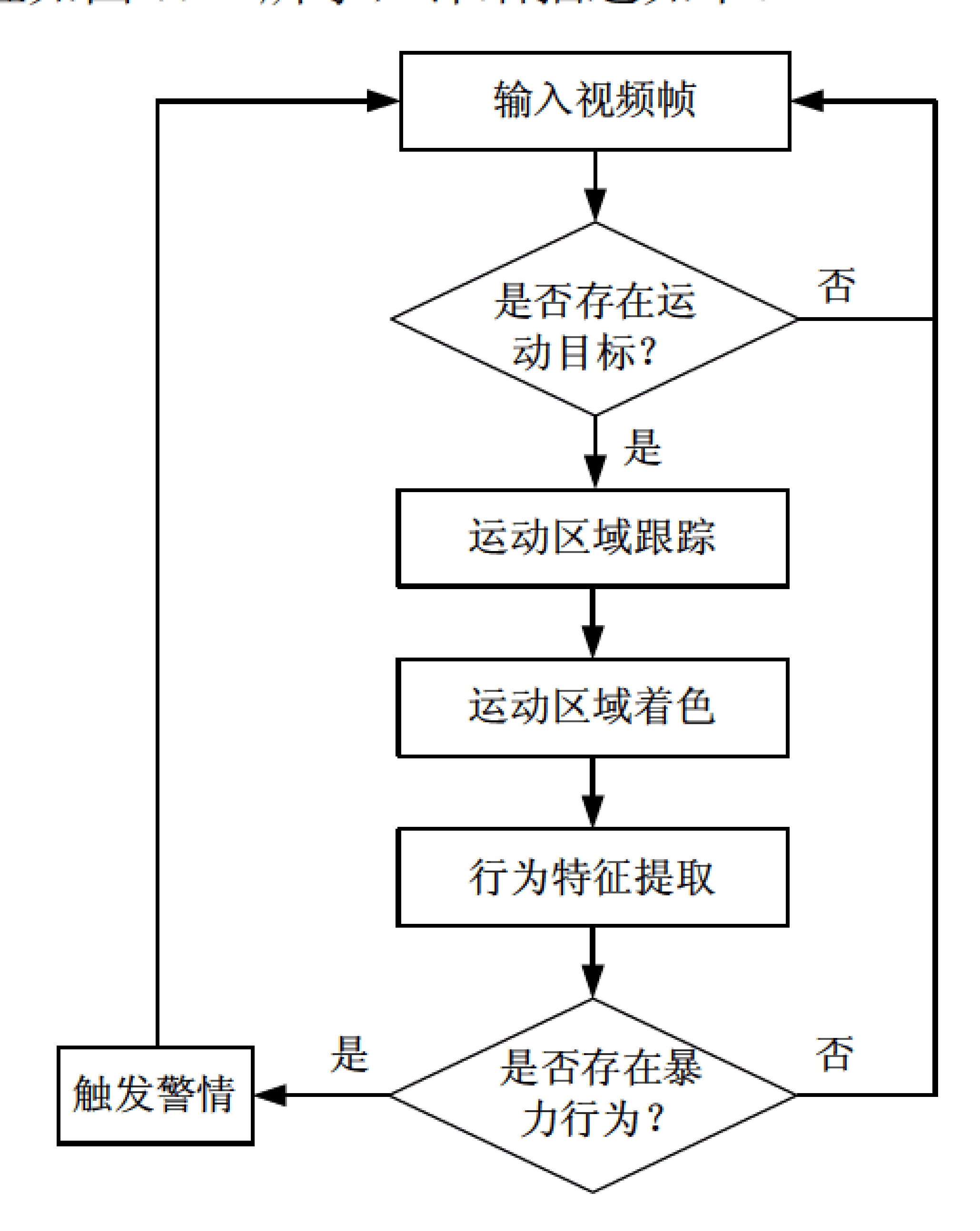


图 7.14　基于运动着色的暴力行为检测系统实现流程

1. 运动区域检测

如 7.3 节所述，暴力行为是由运动目标引起的，因此，首先需要检测场景中是否存在运动目标，只对存在运动目标的场景进行暴力行为检测。采用背景减除法检测运动目标，背景更新采用单高斯模型。对于检测到的所有运动目标，将各目标连通域的最大外接矩形区域记为运动区域 RoI ，得到多个互不重叠的运动区域。

2. 运动区域跟踪

RoI 跟踪策略是：假设 t 时刻图像帧为 F_t，$\tilde{I}_t$ 是由背景估计模块提供的二值匹配分割图像。I_t 是用 F_t 掩膜 $\tilde{I}_t$ 得到的图像。假设 N_t 是在 I_t 中色素点的编号（一个色素点可以是一个人或一群人的一部分），这些色素点通过 $B_i^t \quad (i=1,2,\ldots,N)$ 确定，用其质心 $\vec{C}_i^t=(C_{i,x}^t,C_{i,y}^t)$ 进行表征。色素点集被分成子集 S_k，如果 $\exists B_z \mid d_A(B_i,B_z)<\theta \wedge B_z \in S_k$ ，那么 $B_i \in S_k$，其中 $d_A(,)$ 表示 *Hausdorff* 距离，θ为阈值。S_j 中所有色素点所在的外接矩形区

域记为 $RoI\ R_i$，每个 RoI 的质心由组成它的色素点质心来定义，跟踪器是一个离散函数 $\phi(i):\{1,\ldots,NR_t\}\rightarrow\{1,\ldots,NR_{t-1}\}$，通过利用下一场景中的 RoI 去匹配当前场景中每个 RoI 而得到。

3. 运动区域着色

为获取场景中暴力行为引起的变化特征，引入色彩架构的概念，色彩架构是由 m 个二进制图像 $J_{i,k}{}^t$ $(k=1,2,\ldots m)$ 组成，它和 $RoI\ \ R_i$ 相联系的，可以通过颜色聚类算法对 $RoI\ \ R_i$ 进行色素点分割而得到。考虑到计算效率，一般选用 *CIE Lab* 颜色空间进行颜色采样和聚类。颜色标记从 1 开始，按步长为 1 逐级扫描 *CIE Lab* 颜色空间。为构造二进制图像 $J_{i,k}{}^t$，通过为 F_t 的每个像素分配与它相对应颜色分类标记，去创建一幅新图像 F_t^R。而后，每个图像 $J_{i,k}{}^t$ 通过下列公式得到：

$$J_{i,k}{}^t(x,y)=\begin{cases}1 & if\ F_t^R(x,y)=k\\ 0 & if\ F_t^R(x,y)\neq k\end{cases}$$

其中，(x,y)表示像素坐标，k 是从 1 到 m 的整数。显然，每个 $J_{i,k}{}^t$ 包含了那些 t 时刻颜色为 k 的 RoI 的像素。每个图像 $J_{i,k}{}^t$ 由确定的色素点编号 $n_{i,k}{}^t$ 组成。在 $RoI\ \ R_i$ 中，当它们的质心可以用 $\vec{c}_{i,k,s}{}^t$ 表示时，这些色素点描述颜色 k 的色彩并通过 $R_{i,k}{}^t$ $(s=1,\ldots,n_{i,k}{}^t)$表示。图 7.15 显示部分运动的着色图像。

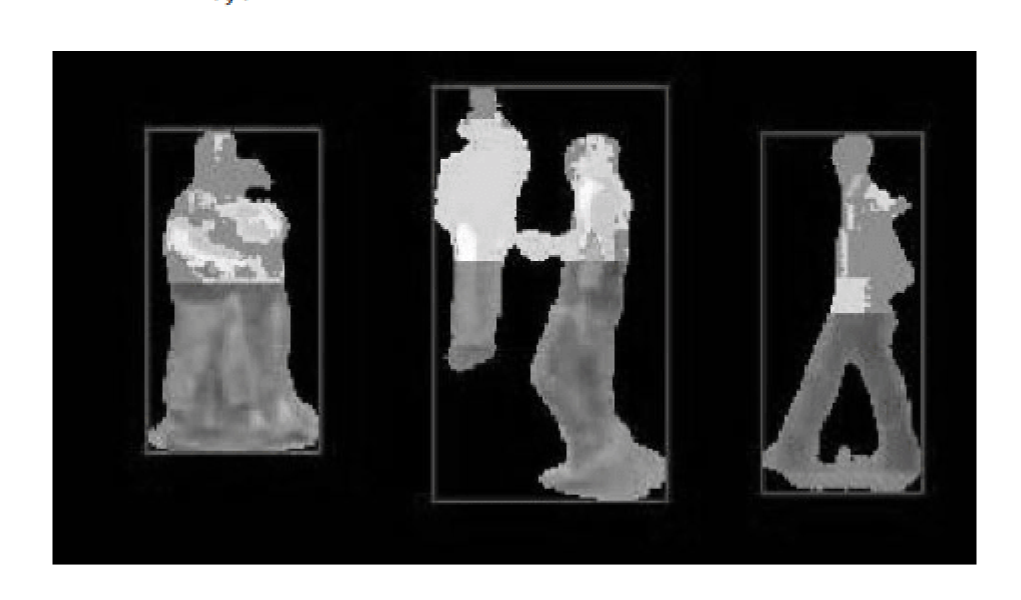



图 7.15　CIE Lab 颜色空间的运动着色图像

4. 行为特征提取

为了分析着色运动，首先估计从 F_{t-1} 帧到 F_t 帧中每个 RoI 的全局运动，在 $\vec{c}_i{}^t$ 和 $\vec{c}_{\phi(i)}{}^t$ 之间建立匹配。每个 RoI 运动区域在时刻 t 的着色运动可以通过匹配 t 时刻 $J_{i,k}{}^t$ 和 t-1 时刻 $J_{\phi(i),k}{}^{t-1}$ 中的色素点进行估计，然后通过函数 $\psi(p,q):\{1,\cdots,n_{i,k}{}^t\}\times\{1,\cdots,n_{\phi(i),k}{}^{t-1}\}\rightarrow\{0,1\}$ 进行着色跟踪，其定义如下：

$$\psi(p,q)=\begin{cases}1 & ,R_{i,k,p}{}^{t}\text{ 与 }R_{i,k,q}{}^{t-1}\text{匹配}\\0 & ，\text{其他}\end{cases}$$

为了描述着色形态的时空复杂度，引入了一个称为总体相对转换能量（$TWE_{i,k}{}^{t}$）的综合性指标，对 t 时刻属于 $RoI\ R_i$ 的 k 类颜色进行着色统计。

首先定义 t 时刻 k 类颜色的变色能量：

$$WE_{i,k}{}^{t}(p,q)=\left\|(\vec{C}_{i,k,p}{}^{t}-\vec{C}_{i}{}^{t})-(\vec{C}_{\phi(i),k,p}{}^{t-1}-\vec{C}_{\phi(i)}{}^{t})\right\|^{2}$$

k 类颜色变色能量总计定义如下：

$$TWE_{i,k}{}^{t}=\sum_{p=1}^{n_{i,k}{}^{t}}\sum_{q=1}^{n_{\phi(i),k}{}^{t-1}}\psi(p,q)\cdot WE_{i,k}{}^{t}(p,q)$$

因为共有 m 类不同的颜色，因此将有 m 个 TWE 使用最大值操作进行融合：

$$MWE_i^t=\max_{k\in[1,m]}\{TWE_{i,k}{}^{t}\}$$

5. 暴力行为判定

如果 t 时刻 $RoI\ R_i$ 的 MWE_i^t 超过预先确定的阈值，则认为区域 $RoI\ R_i$ 发生暴力行为，并认定该时刻的图像帧为暴力帧。当一定时间内存在多帧图像为暴力帧时，则判定当前场景发生暴力行为，触发警情信号。

第 8 章

可疑行为检测系统

与暴力行为相比，可疑行为不直接威胁公共安全，但是其潜在危害很大，甚至比暴力行为更大。对视频监视而言，可疑行为的发生概率远远大于暴力行为，是社会治安、反恐和维稳的重点。

8.1 可疑行为

可疑行为是指可能侵害公民的人身和财产安全，甚至引发重大公共安全事件的行为，可疑行为多种多样，与场景上下文和应用环境息息相关，譬如可能破坏社会安定的街头群体聚集行为、可能危害他人安全的尾随行为等。

按照参与人员数目的不同，可疑行为可分为基于单人行为的徘徊、奔跑、躬身、匍匐、跳跃、下蹲、倒地、越界、攀爬、遗留物品等，基于两人行为的尾随等，基于多人行为的聚集等。

与暴力行为不同，可疑行为的激烈程度小，危险性不明确，要根据特定场合和上下文进行推断。图 8.1 列举了几种关注度较高的可疑行为，其特点如表 8.1 所示。

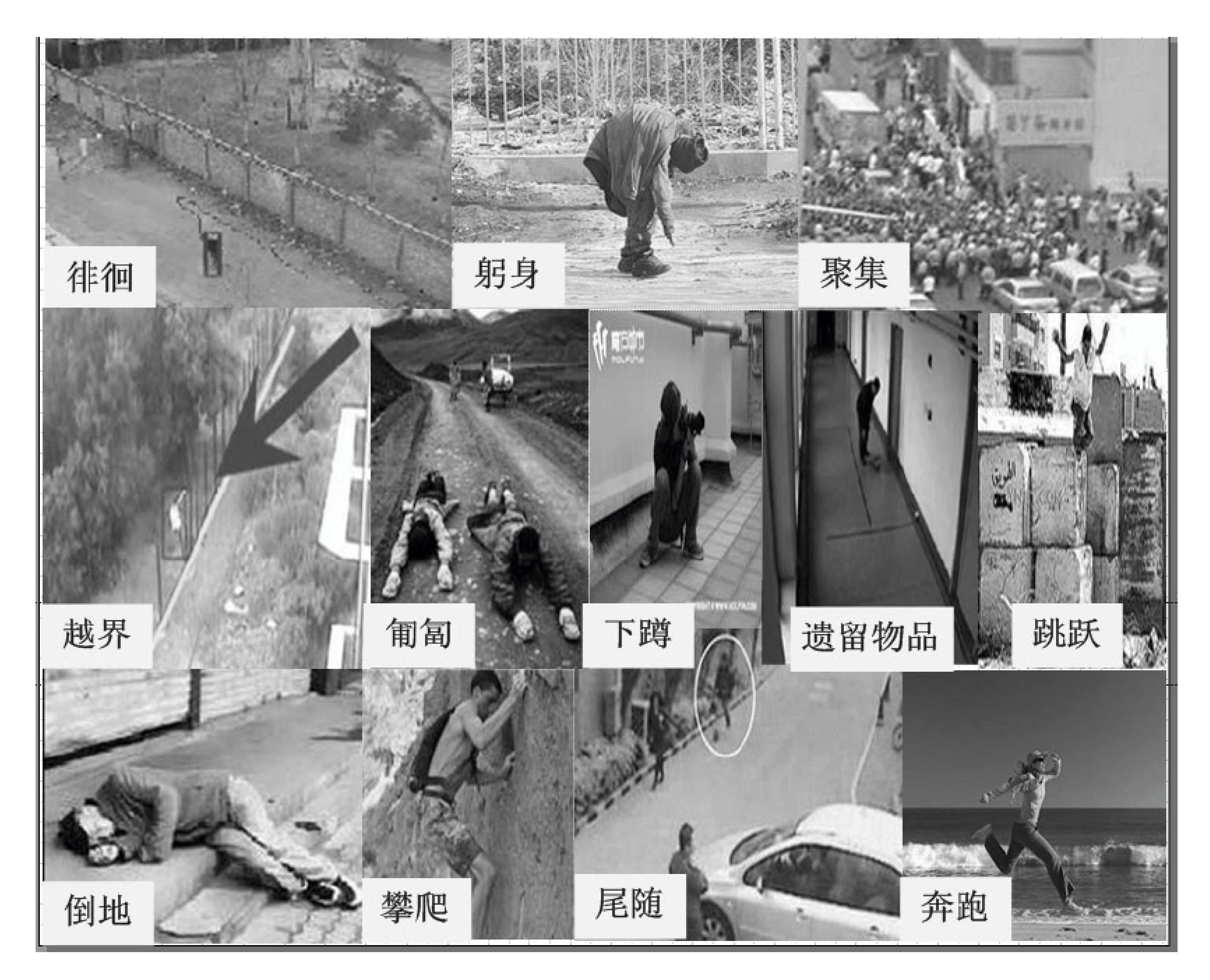


图 8.1　常见可疑行为示例

表 8.1　常见可疑行为特点

行为	特点
徘徊	人体运动轨迹在空间上有较大重复，在时间上有明显周期性
奔跑	人体移动速度很快
躬身	人体运动时呈弯腰姿态
匍匐	人体运动时呈卧地前行姿态
跳跃	人体在垂直方向运动，移动速度快，且运动过程中身体做伸缩运动
下蹲	人体从直立姿态变为蹲着姿态
倒地	人体从直立姿态变为横躺姿态
越界	人体运动轨迹进入某虚拟周界，且运动方向朝向虚拟周界
遗留	人体和随身携带物品在同向运动过程中，物品和人体突然分离，物品停止运动，人体继续运动
攀爬	人体在垂直方向向上运动，四肢呈伸展姿态且呈周期性运动
尾随	两人运动轨迹基本一致，且相对距离变化不大
聚集	场景中人体目标的运动从有序变为混乱，且某区域人口密度增加，人与人之间的距离减小

8.2　可疑行为检测

可疑行为检测包括用户交互模块、视频采集模块、可疑行为检测模块、视频编码与网络传输模块、数据存储与显示模块等，其核心是可疑行为检测模块。这些模块在第 7 章有过详细介绍，主要区别为可疑行为理解模块。

与暴力行为不同，可疑行为没有剧烈运动的典型性特征，且可疑行为种类繁多，不同的可疑行为可能具有截然不同的行为特征，譬如多人的群体聚集行为和单人的倒地行为差异很大，单人的徘徊行为和单人的倒地行为也有明显差异。因此，可疑行为的理解没有通用方法，需要针对一种或者几种类似的可疑行为展开算法研究。

如何选择充分的特征有效表征可疑行为是行为理解的关键，特征选择要考虑的问题主要有两个：一是目标显著与稳健特征的选择，二是目标特征的精确测量。待选择的目标特征应具备如下特点。

- 可靠性：同类目标的特征值相似。
- 可区分性：不同类目标的特征值具有明显差异，目标与背景的特征值也具有明显差异。
- 独立性：同一目标中，各特征相互独立，互不相关。
- 精简性：原始特征通过映射或变换方法进行降维。

描述可疑行为的特征主要是运动特征和形状特征，譬如采用运动特征中的轨迹可以很好地描述徘徊和尾随行为；采用运动特征中的速度可以很好地描述跳跃和奔跑行为；采用运动特征中的运动方向可以较好地描述越界和倒地行为；采用形状特征可以较好地区分人体和非人体，从而描述聚集、遗留物品等可疑行为；采用形状特征中的姿态特征可以较好地描述匍匐、躬身等可疑行为。

在实际应用中，需要首先深入分析可疑行为的一些先验知识，然后选择具有典型性和稳健性的行为特征。

可疑行为检测系统的评价指标与暴力行为检测系统相同，详见第 7 章的内容。可疑行为数据库一般从第 7 章中表 7.2 所述的行为数据库中抽取，这里不再赘述。

8.3　基于轨迹特征的可疑行为检测系统

人体运动的轨迹特征常被用于检测可疑行为，张瑞玉、张锦等人利用人体运动轨迹特征检测徘徊行为，胡卫明等人利用轨迹特征检测停车场中是否有可疑人员等。事实上，轨迹特征还可以用于检测多种可疑行为，谢剑斌等人提出一种基于轨迹特征的可疑行为

检测系统，利用轨迹特征检测徘徊、奔跑、匍匐、倾倒、躬身等可疑行为。

8.3.1　系统结构

基于轨迹特征的可疑行为检测流程如图 8.2 所示，该系统主要包括人体目标检测、轨迹建模、特征提取与分类 3 个模块。

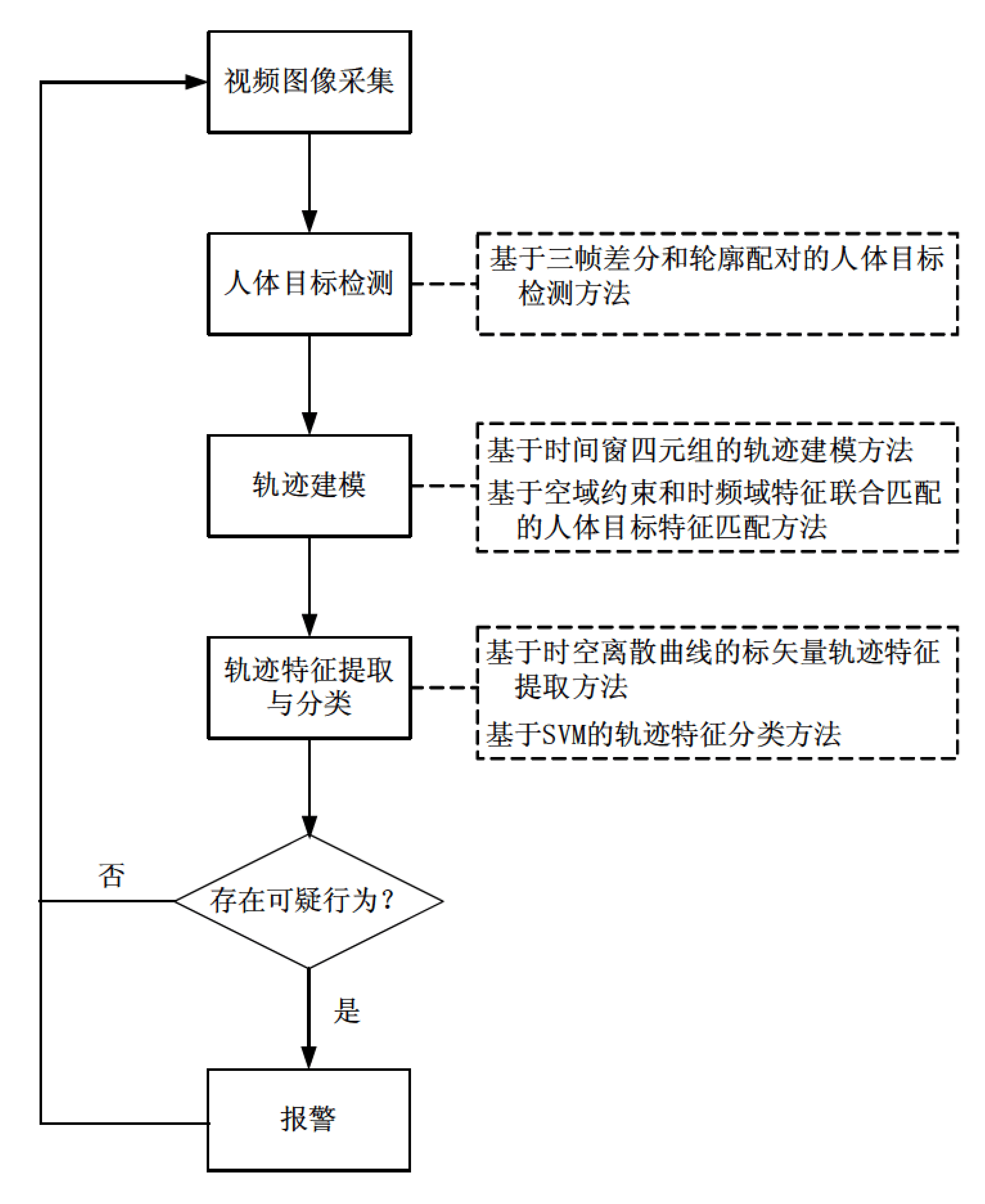


图 8.2　基于轨迹特征的可疑行为检测流程

8.3.2　人体目标检测

在监视场景中，重要的目标是运动的人体目标，首先采用三帧差分法检测运动目标，然后采用轮廓配对法筛选人体目标。其中，三帧差分法的伪代码如下。

算法 8.1 三帧差分法

输入：三帧图像 I_1、I_2、I_3。

过程：1.计算帧差图像 E_1、E_2：

$$E_1 = |I_3 - I_1|$$

$$E_2 = |I_3 - I_2|$$

2.计算帧差图像均值，乘以加权系数，作为自适应阈值 T：

$$m = \frac{1}{2 \times M \times N} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [E_1(i,j) + E_2(i,j)]$$

$$T = \beta \times m$$

其中，$W \times H$ 为视频图像尺寸，β 为加权系数，β 取 10。

3.阈值分割，得到二值图像 MR。

$$MR(i,j) = \begin{cases} 1 & ,E_1(i,j) \geq T \quad 且 \quad E_2(i,j) \geq T \\ 0 & ,otherwise \end{cases}$$

MR 中数值为 1 的像素点记为运动目标点。

输出：前景与背景分离的二值图像。

二值图像 MR 中的目标难免出现断裂和“孔洞”现象，同时存在噪声。为此，首先采用中值滤波方法平滑目标块，去除噪声；然后采用数学形态学的开运算操作填补目标块的“孔洞”，合并相邻的目标块；最后采用 8-邻接连通方法搜索和标记目标。

由于运动目标并不一定都是人体，因此需要对目标的属性进行判断，尽可能多地剔除动物、车辆等干扰目标，降低虚警率。采用轮廓配对方法剔除干扰目标，伪代码见算法 8.2。

算法 8.2 轮廓配对方法

输入：运动目标块。

过程：1.检测目标块的轮廓，轮廓点(x,y)满足两个条件：

条件 1：$MR(x,y)=1$。

条件 2：$MR(x,y+1)+MR(x,y-1)=1$，

或 $MR(x+1,y+1)+MR(x-1,y-1)=1$，

或 $MR(x+1,y-1)+MR(x-1,y+1)=1$，

或 $MR(x+1,y)+MR(x-1,y)=1$。

2. 采用归一化傅立叶描述子表示目标轮廓。

对坐标为(x, y)的第 n 个轮廓点，记 $X[n]=x$，$Y[n]=y$，计算傅立叶描述子：

$$a(u)=\sum_{k=0}^{K-1}(X(k)+jY(k))e^{-j2\pi uk/K}, u=0,1,\ldots,K-1$$

其中，K 为轮廓点总数。由于傅立叶描述子与形状尺度、方向和曲线起始点有关，故需进行归一化：

$$d(u)=\frac{a(u)}{a(1)}, u=1,2,\ldots,K-1$$

3. 采用欧式距离进行轮廓配对，判断目标属性。假设待识别目标的傅立叶描述子为 $d_1(u)$，人体目标的傅立叶描述子为 $d_2(u)$，则二者的形状差异为：

$$d=\sqrt{\sum_{u=1}^{K-1}\|d_1(u)-d_2(u)\|^2}$$

设定固定阈值 D，这里取 $D=0.02$。如果 $d<D$，则认为该目标为人体目标，否则认为该目标为干扰目标。

输出：运动目标块是否为人体目标的结论。

8.3.3　轨迹建模

人体的轨迹信息是判断人体可疑行为的重要依据之一，如何建立稳定可靠的人体轨迹模型是判别人体可疑行为的基础。

1. 轨迹四元组

采用基于时间窗四元组的轨迹建模方法，轨迹的四元组记为：

$$TR=\{i,f,P(x,y),d(u)\}$$

其中，i 表示目标序号，f 表示视频帧号，$P(x,y)$表示目标质心坐标，$d(u)$表示目标轮廓描述子。

对于每一帧二值图像 MR 中筛选的某人体目标，依次记录视频帧号、目标质心坐标、目标轮廓描述子信息。其中，轮廓描述子信息在 8.3.2 节已经求得，目标质心坐标为：

$$\bar{x} = [\sum_{x=0}^{W-1}\sum_{y=0}^{H-1} xMR(x,y)]/[\sum_{x=0}^{W-1}\sum_{y=0}^{H-1} MR(x,y)]$$

$$\bar{y} = [\sum_{x=0}^{W-1}\sum_{y=0}^{H-1} yMR(x,y)]/[\sum_{x=0}^{W-1}\sum_{y=0}^{H-1} MR(x,y)]$$

2. 目标序号的标记方法

对于第一帧视频图像，依次标记各个人体目标的序号。对于后续视频图像中出现的各个人体目标，和前一帧中的各个人体目标进行特征匹配，如果匹配成功，则该目标序号标记为前一帧相匹配的目标序号；否则，该目标标记为新的序号。

3. 人体目标特征匹配

采用基于空域约束和时频域特征联合匹配方法实现人体目标特征匹配，具体步骤如下。

步骤01　空域约束

一般地，即使是人体快速奔跑的速度也不可能达到视频实时采样速度，相邻两帧视频图像中同一人体目标的轮廓是有重叠的，因此，依据空域约束区分明显不是同一个人体的目标。假设重叠点为(x, y)，则在前后两帧二值图像中，(x, y)必须满足两个条件。

条件 1：$MR_1(x,y)=1$

条件 2：$MR_0(x,y)=1$

其中，MR_1表示当前帧目标块，MR_0表示前一帧目标块。

如果前后两帧人体目标有重叠点，则认为两个人体目标有可能是同一个目标，继续下一步匹配；否则，认为两个人体目标不匹配，终止目标匹配过程。

步骤02　频域特征匹配

可采用傅立叶描述子特征进行频域特征匹配，假设当前帧目标的傅立叶描述子为$d_1(u)$，前一帧目标的傅立叶描述子为$d_2(u)$，则目标之间的频域特征差异为：

$$d = \sqrt{\sum_{u=1}^{K-1} \|d_1(u) - d_2(u)\|^2}$$

设定固定阈值D_2，$D_2<D$，这里取$D_2=0.013$。如果$d<D_2$，则认为前后两帧人体目标有可能是同一个目标，继续下一步匹配；否则，认为两个人体目标不匹配，终止目标匹配过程。

步骤03　时域特征匹配

下面采用梯度向量特征进行时域特征匹配。

按照梯度算子计算各像素点梯度：

$$G_x(i,j) = I(i+1,j) - I(i-1,j)$$

$$G_y(i,j) = I(i,j+1) - I(i,j-1)$$

梯度模值为：

$$G(i,j) = \sqrt{G_x(i,j)^2 + G_y(i,j)^2}$$

梯度方向为：

$$\alpha(i,j) = \tan^{-1}(G_x(i,j)^2 / G_y(i,j)^2)$$

把[-π/2,π/2]的梯度方向均匀划分为 9 个区间（记为 $area_k$,$1\le k\le 9$），则各个像素点在分量区间上的 9 维梯度向量特征为：

$$V_k(i,j) = \begin{cases} G(i,j) & ,\alpha(i,j) \in area_k \\ 0 & ,otherwise \end{cases}$$

目标块的平均梯度向量特征为：

$$V_k = \frac{1}{W \times H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} V_k(i,j), 1 \le k \le 9$$

假设当前帧目标的梯度向量特征为 V_1，前一帧目标的梯度向量特征为 V_2，则目标之间的时域特征差异为：

$$v = 1.0 - \exp(-\sqrt{\sum_{k=1}^{9} (V_{1k} - V_{2k})^2} / 10000)$$

设定固定阈值 D_3，这里取 D_3=0.14。如果 $v<D_3$，则认为前后两帧人体目标是同一个目标；否则，认为两个人体目标不匹配。

由于人体目标的轨迹与时间有关，于是在得到每一帧视频图像中各个目标的四元组之后，采用时间窗法得到时间窗四元组，记为 W_{TR}：

$$W_{TR} = \{TR_k \mid t_0 \le k < t_0 + t_d\}$$

其中，t_0 表示起始视频帧号，t_d 表示时间窗宽度，即间隔帧数。

8.3.4 轨迹特征提取

在获取各个人体目标的时间窗四元组之后，可提取轨迹特征。

将时间窗四元组中的质心坐标相连接，可以得到一条时空离散曲线，如图 8.3 所示。该曲线反映人体目标在时间窗内的运动轨迹，是辨别可疑行为的重要依据。采用基于时空离散曲线提取轨迹的标量特征和矢量特征，实现步骤如下。

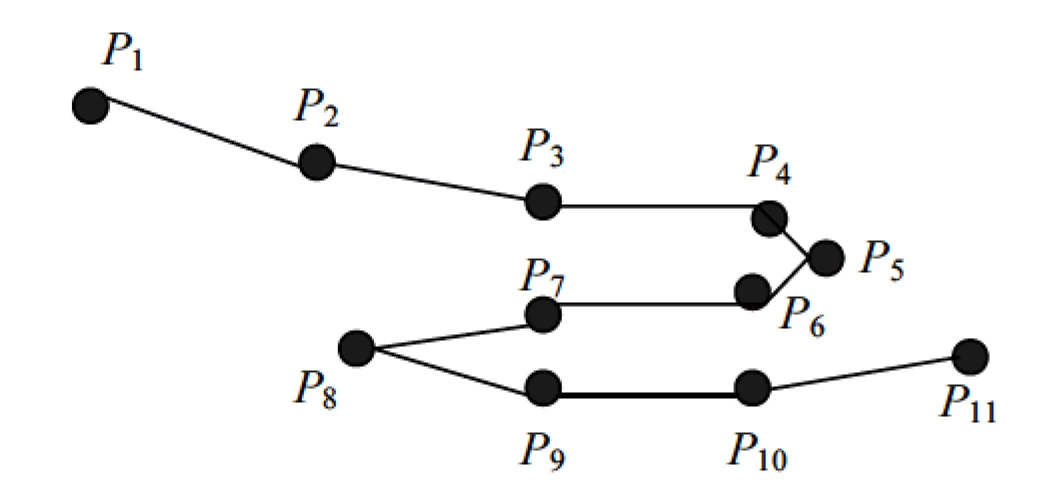


图 8.3 时空离散曲线

1. 轨迹的标量特征提取

时空离散曲线标量特征包括：广义曲率、时空长度和时空拐点数目。

❑ 广义曲率

首先，计算时空离散曲线上各离散点与相邻两点的夹角，作为该离散点的角度特征，以 P_2 点为例，其角度特征为：

$$\alpha_2 = \arccos[(\overline{P_2P_3} + \overline{P_2P_1} - \overline{P_1P_3})/(2\sqrt{\overline{P_2P_3} \times \overline{P_2P_1}})]$$

其中

$$\overline{P_2P_3} = (P_{2x} - P_{3x})^2 + (P_{2y} - P_{3y})^2$$
$$\overline{P_2P_1} = (P_{2x} - P_{1x})^2 + (P_{2y} - P_{1y})^2$$
$$\overline{P_1P_3} = (P_{1x} - P_{3x})^2 + (P_{1y} - P_{3y})^2$$

然后，取所有离散点角度特征的平均值，作为广义曲率。这里以图 8.3 的时空离散曲线为例，取中间 9 个点的角度均值，得到的广义曲率为：

$$\overline{\alpha} = \sum_{k=2}^{9} \alpha_k$$

❑ 时空长度

时空长度特征可以由时空离散曲线上的离散点数目代替，即为时空离散曲线上的离

散点数目。对于图 8.3 所示的时空离散曲线，其时空长度特征为 11。

❑ 时空拐点数目

在时空离散曲线上，当离散点的角度特征小于 $\pi/2$ 时，认为该点为时空拐点。在时空离散曲线上，时空拐点的个数为时空拐点数目。在图 8.3 所示的时空离散曲线上，P_5、P_8 为时空拐点，时空拐点数目为 2。

2. 轨迹的矢量特征提取

对于时空离散曲线上的每一个离散点，提取空域和时域两个矢量特征。

❑ 空域矢量

空域矢量描述人体目标在行进过程中的身体倾向，用来区分人体是直立姿态还是前倾或后倾等倾倒姿态，甚至完全倒地姿态，有助于辨别人体奔跑、倾倒、匍匐、躬身等行为。

空域矢量的获取方法是：首先由四元组的轮廓描述子恢复人体轮廓形状；然后采用椭圆曲线拟合方法获取人体椭圆形状；最后提取椭圆的长轴矢量，作为空域矢量。

❑ 时域矢量

时域矢量描述人体目标在行进过程中的运动情况，对于时空离散曲线上的任一点，其时域矢量的模值为该点与下一点的欧式距离，时域矢量的方向为该点指向下一点的方向和水平方向的夹角。

以图 8.3 中 P_2 点为例，其时域矢量的模值为：

$$\tau = \sqrt{(P_{2x} - P_{3x})^2 + (P_{2y} - P_{3y})^2}$$

方向为：

$$\theta = \arctan[(P_{2x} - P_{3x})/(P_{2y} - P_{3y})]$$

8.3.5 轨迹特征分类

由于行为特征随机性强，难以依据模板匹配或最小距离等方法进行分类。可采用 SVM 方法进行行为特征分类。SVM 是在统计学习理论基础上发展起来的一种学习方法，可以有效解决小样本、模型选择和非线性问题，具有很强的泛化性能。核函数是 SVM 算法的关键，选择径向基函数作为 SVM 的核函数：

$$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right)$$

在训练阶段，首先选择尽可能多的正负样本，正样本为包含徘徊、奔跑、匍匐、倾倒、躬身等行为的视频，负样本为包含正常行走、聚集、聊天等行为的视频；然后采用轨迹四元组方法建立轨迹模型，采用基于时空离散曲线提取轨迹的标量特征和矢量特征；最后采用 SVM 方法进行训练，得到分类器。

在识别阶段，首先对实时视频中的各个人体目标建立轨迹模型，提取轨迹特征；然后将轨迹特征输入由训练阶段得到的分类器进行分类；最后判别待检测视频中是否存在可疑行为，如果存在可疑行为，则启动声光报警。

8.4 基于运动方向的可疑行为检测系统

运动方向是可疑行为检测的重要依据之一，胡芝兰等人提出了一种基于运动方向的可疑行为检测系统，利用人体的运动方向特征检测聊天、病倒、放包、取包、在门附近徘徊以及进出门等可疑行为。

8.4.1 系统流程

基于运动方向的可疑行为检测系统流程，如图 8.4 所示，该系统包括背景边缘模型、前景帧判断、行为特征描述和 SVM 分类器等模块。

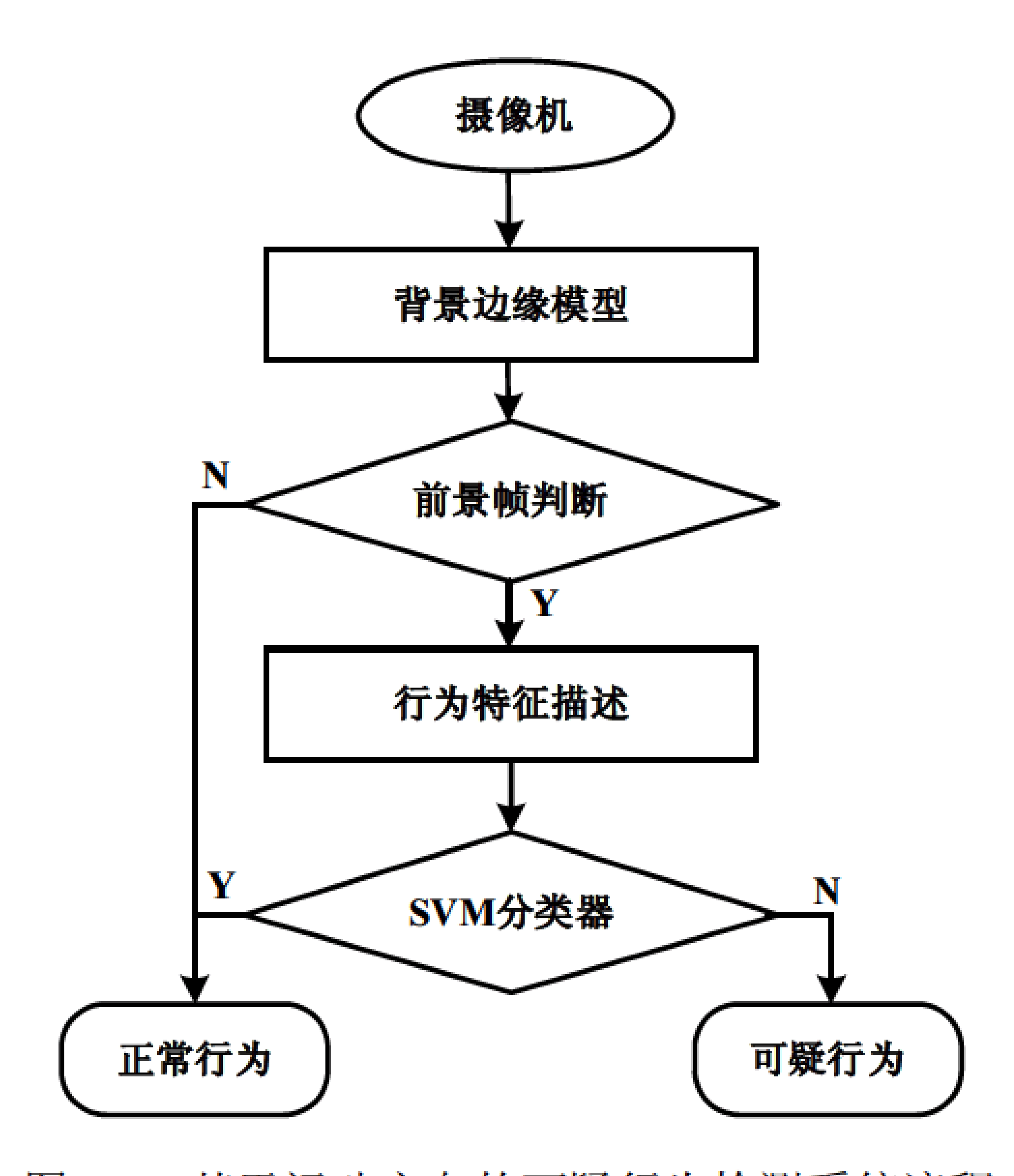


图 8.4 基于运动方向的可疑行为检测系统流程

8.4.2 背景边缘模型

传统前景检测方法提取干净的前景区域，本系统前景检测只是为了判断当前帧是否为前景帧。由于边缘对光照变化的鲁棒性要比区域或者像素强很多，且背景的边缘位置相对固定，因此考虑构建背景边缘模型，通过区分前景边缘和背景边缘，实现复杂光照环境下的前景帧检测。具体实现方法是：在某段时间内，统计视频帧各像素点出现边缘的概率，构建背景边缘模型。记 $P_b(i,j,t)$为像素点(i,j)在当前帧 t 为背景边缘的概率：

$$P_b(i,j,t)=\sum_{k=t-T}^{t-1}\frac{E(i,j,k)}{T}, \quad i=1,2,\cdots,W, j=1,2,\cdots,H$$

其中，$E(.,.,k)$为第 k 帧所对应的边缘图像，该边缘图像通过 Canny 算子检测得到；W 和 H 分别为每帧图像的宽和高；T 为背景边缘模型的更新时间。

对于前景目标，即使相对静止的行为（如站立），同一姿势的滞留时间也不会很长，即前景目标的轮廓边缘仍处于小幅度运动中，所以 T 的取值不需很大，以便及时适应背景边缘的更新。根据实验分析，更新时间 T 取 1500 帧，即为 60s（25fps）。

8.4.3 前景帧判断

采用 Canny 算子，可以得到当前帧 t 的边缘图像，记为 $E(.,.,t)$；依据背景边缘模型，判断边缘点(i,j)是否为前景点：

$$F(i,j,t)=\begin{cases}1 & E(i,j,t)=1 \text{ 且 } P_b(i,j,t)<TB\\ 0 & \text{其他}\end{cases}$$

如果背景边缘概率低于阈值 TB，则该边缘点为前景边缘点，否则为背景边缘点。阈值 TB 的设置对检测结果有较大影响，阈值设置太小时前景边缘点易漏检，阈值设置太大时背景边缘点会由于噪声干扰或相机抖动影响而误检。实验中取 TB=0.2。

由于光照以及灰尘、噪声等影响，在前景边缘图中会存在一定的噪声边缘点。一般地，在背景帧图像的前景边缘图中，噪声边缘点少且分布散；而在前景帧的前景边缘图中，边缘点多且分布相对集中。因此，可以根据前景边缘点的数量和分布情况区分前景帧和背景帧，方法是：首先，依据邻域信息去除噪声点，具体地，如果某前景边缘点所在 8 邻域内的边缘点少于 3 个，则认为该 8 邻域内全部为噪声点；然后，统计图像中前景边缘点数量，记为 Nt，如果 $Nt\geqslant30$，判定该帧为前景帧，否则，判定为背景帧。

8.4.4　行为特征描述

不同的监视视频中，行为发生的位置和速度千变万化，行为个体的数量和尺寸也各不相同。为了保证行为识别性能，行为描述特征需要具有位置和尺度不变性，且对个体数量和运动速度不敏感。块运动方向可以满足上述要求，图 8.5 显示了不同行为之间块运动方向的差异，其中，用 bin 表示块运动方向归一化直方图，X 轴表示每个 bin 的中心值，Y 轴表示该视频段属于对应 bin 的运动方向的比例；若将打架换为病倒，则对应直方图高度应适当降低。

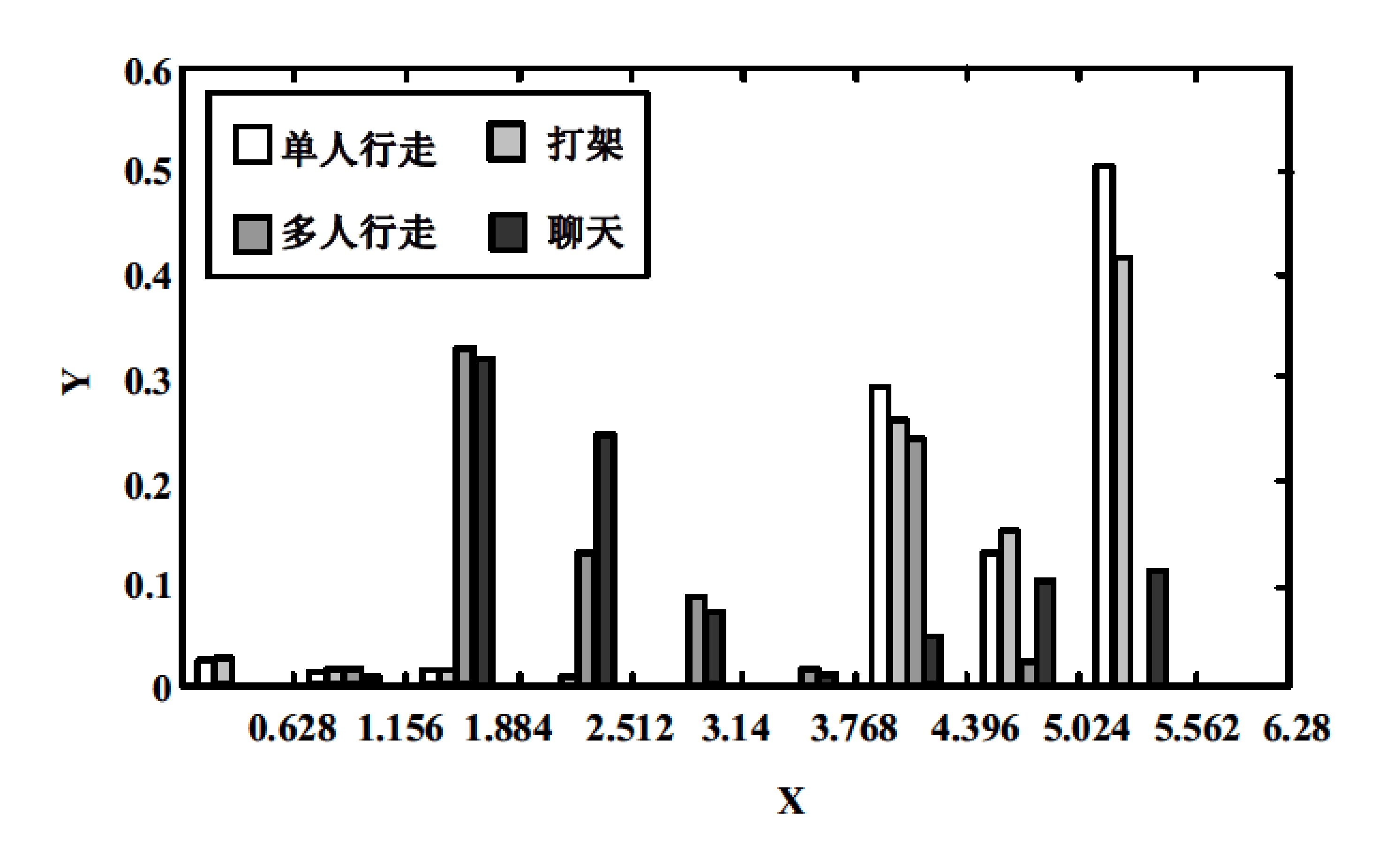


图 8.5　不同行为的运动方向归一化直方图

可见聊天、病倒等可疑行为与正常行走行为存在明显差异，因此块运动方向归一化直方图可以描述行为特征，相关伪代码见算法 8.3。

算法 8.3　行为特征描述

输入：视频流。

过程：1.视频分段。等间隔抽取 10 帧图像，帧与帧间隔为 10。该视频分段基本对应一个完整的动作。

2. 提取视频分段中的所有前景帧，如果该视频段中前景帧的比例小于 80%，认为该视频段不存在前景帧，块运动方向归一化直方图数值全为零，退出；否则进入下一步。

3. 对于每一前景帧中提取运动幅度不为零的块，计算块运动方向。对于 VGA 视频，块大小取为 8×8。

4. 对该视频段的所有块运动方向进行归一化直方图统计，得到该视频段的行为描

述特征。在直方图统计过程中，采用 13 个直方块(bin)，每个 bin 的中心值在[0,2π]之间均匀求出。

输出：块运动方向归一化直方图。

8.4.5　SVM 分类

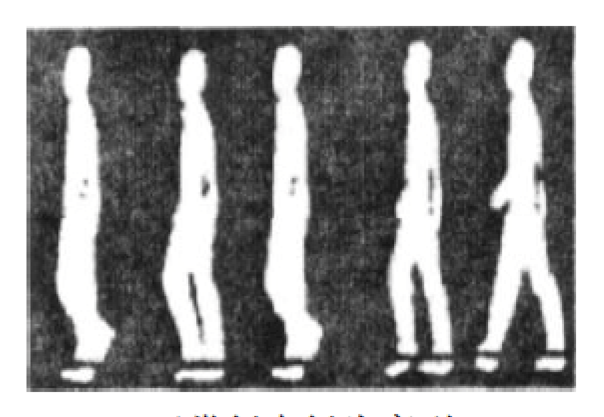
对于提取到的行为描述特征，将正常行走行为的特征作为正样本，将聊天、病倒、放包、取包、在门附近徘徊以及进出门等可疑行为的特征作为负样本，采用 SVM 分类器进行训练和分类。

SVM 分类器在前文已经介绍，这里不再赘述。

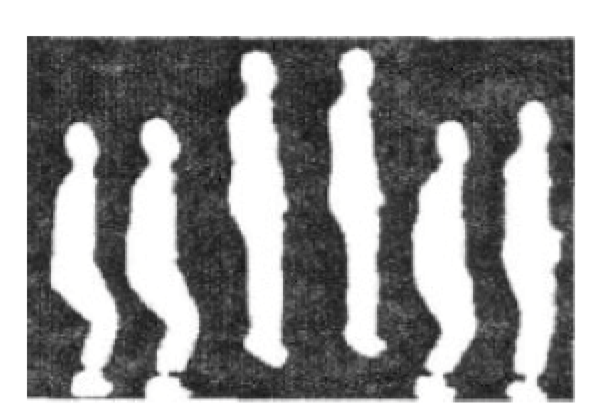
在经过分类器分类之后，如果是可疑行为，则触发警情信号。

8.5　基于形状特征的可疑行为检测系统

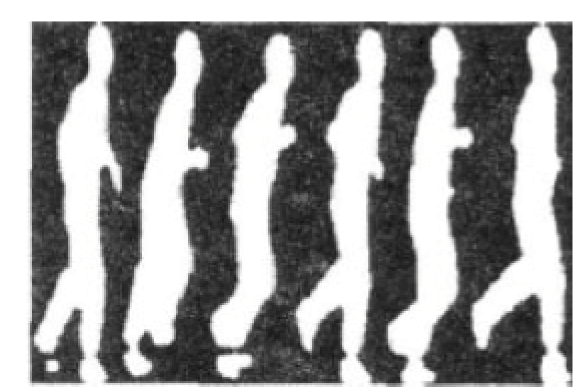
对于跳跃、奔跑、倒地、下蹲、挥手和手拿异物等可疑行为，在行为发生时，人体目标的形状有显著变化，如图 8.6 所示。因此，形状特征是检测可疑行为的重要依据之一。



正常行走行为序列



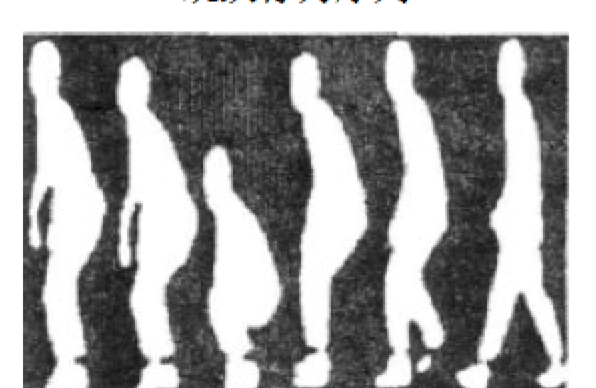
跳跃行为序列



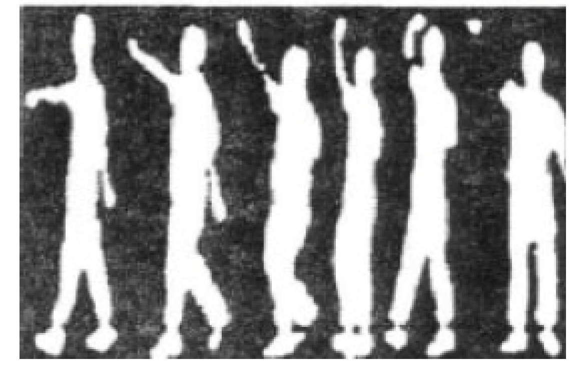
奔跑行为序列



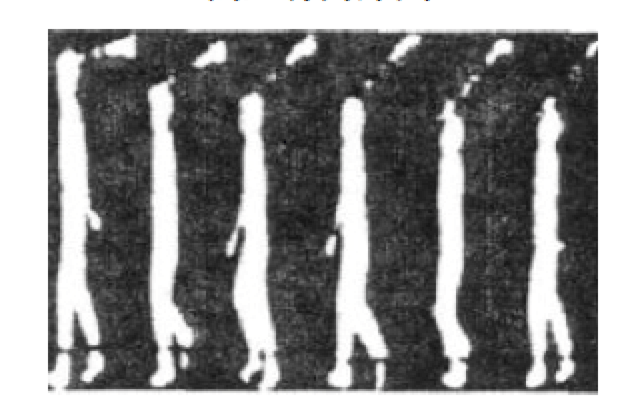
倒地行为序列



下蹲行为序列



挥手行为序列



手持异物行为序列

图 8.6　行为视频序列

印勇等人设计了一种基于形状特征的可疑行为检测系统，用于检测跳跃、奔跑、倒地、下蹲、挥手和手拿异物等可疑行为。该系统首先采用背景差分法提取运动人体目标；然后采用 Hu 矩特征描述人体目标的形状；最后采用 SVM 对形状特征进行训练和分类，检测可疑行为。

运动人体目标提取的方法和 SVM 分类器在前文已有论述，这里主要介绍形状特征的提取方法。

Hu 矩特征是描述目标形状的常用方法，对平移、旋转、尺度具有不变性。视频图像的 7 个 Hu 不变矩特征计算方法为：

$$M_1 = \eta_{20} + \eta_{02}$$

$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}{}^2$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$M_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} + \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$M_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$M_7 = (3\eta_{12} - \eta_{30})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{12} + \eta_{30})^2]$$

其中：

$$\eta_{pq} = [\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(i-\bar{i})^p (j-\bar{j})^q I(i,j)] / \{[\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} I(i,j)]^{\frac{(p+q)}{2}+1}\}$$

$$\bar{i} = [\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} iI(i,j)] / [\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} I(i,j)]$$

$$\bar{j} = [\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} jI(i,j)] / [\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} I(i,j)]$$

$I(i,j)$表示像素点(i,j)处的亮度，M 和 N 分别表示图像的宽度和高度。

对于跳跃、奔跑、倒地、下蹲、挥手和手拿异物等可疑行为，一般持续时间为 1~2s。为减少数据量，采用间隔抽取视频方法提取形状特征。这里取 3 帧为一个间隔，共提取

15 帧图像为一个样本。对于每一帧图像，提取的 Hu 矩特征可以用向量表示为：

$$\Phi_j = \{M_1, M_2, M_3, M_4, M_5, M_6, M_7\}, \quad j = 1, 2, \cdots, 15$$

行为序列的形状特征可以用一个 15×7 的集合表示：

$$A = [\Phi_1 \quad \Phi_2 \quad \cdots \quad \Phi_{15}]^T$$

对于提取到的形状特征，采用 SVM 分类器进行训练和分类，实现可疑行为的检测。如果检测到可疑行为，则触发警情信号。

第 9 章

海量视频摘要系统

随着多媒体技术的迅猛发展和视频采集设备的普及，视频资源飞速膨胀，海量视频数据中存在巨大冗余，严重影响后续的视频分析与检索效率。可采用视觉计算、机器学习、人工智能等方法，从海量视频数据中，自动提取有价值的视频画面，以降低冗余度，形成视频摘要。

9.1 视频摘要

在分析视频数据时，将相关主题的多个视频搜索结果进行整合和精简，按照某种逻辑关系以直观形式展示，即视频摘要（Video Abstraction），视频摘要可以提供简洁、准确、全面的视频信息，提高海量视频数据的分析效率。

视频摘要有多种媒体和表现形式，可以是一段文字、一幅图像、一段视频，或者由多种媒体组合而成。

视频摘要由多媒体内容分析与检索（Multimedia Content Analysis and Retrieva1）、多媒体搜索排序（Multimedia Search Ranking）、近似重复检测（Near Duplicate Detection）等相关技术发展而来，研究手段从底层语义分析到高层语义分析，处理对象从单个视频、多个视频到海量视频。

依据表现形式的不同，视频摘要可分为静态摘要和动态摘要。

1. 静态摘要

静态摘要主要有 4 种形式。

- 标题（Title）

标题是对视频进行简短描述的一个词或一句话，采用简单方式表现视频内容。

标题简单便捷，但是传达的信息量比较少。

- 海报（Poster）

海报是指从原始视频中抽取的某一帧或几帧关键的图像，有时还配有相关的文字信息，也叫视频缩略图，或者视频代表帧。

海报可以给用户直观感受，但是仅能表现某些时刻的视频画面，很难表示视频的具体内容和发生的事件。

- 故事板（Storyboard）

故事板是指对视频进行镜头切分以及抽取出所有关键帧之后，将这些关键帧按照时间顺序组合成列表。

故事板包含更多视频语义，提供的视频信息比较完整。

- 幻灯片（S1ide）

幻灯片是指由视频中抽取的部分关键帧组成的 GIF 文件。

幻灯片应用于需要在一个页面中显示尽量多的视频，如视频检索时返回视频列表的显示。

2. 动态摘要

动态摘要是一种缩略视频，由原始视频中抽取的一些分散的镜头拼接而成，这些镜头最能体现原始视频的主题。动态摘要保留了原视频风格，提供给用户的信息丰富。

动态摘要用途广泛，如电影和电视剧的预告片等。

9.2　视频摘要过程

海量视频摘要包含 3 个基本过程。

- 视频结构解析（Video parsing）

这是第一个过程，将视频流按照帧、镜头、场景等层次结构进行分段。

- 特征提取和表示（Feature extraction & representation）

这是第二个过程，将视频中的纹理、颜色、形状和运动等语义信息（或视觉特征）提取出来。

❑ 内容摘要（Content abstraction）

这是第三个过程，从原视频流中提取出一些镜头、场景以及故事情节的子集，代表原视频的内容。

一段视频的典型结构如图 9.1 所示，视频结构解析的目的就是将视频数据拆分为表征不同层次含义的数据单元。

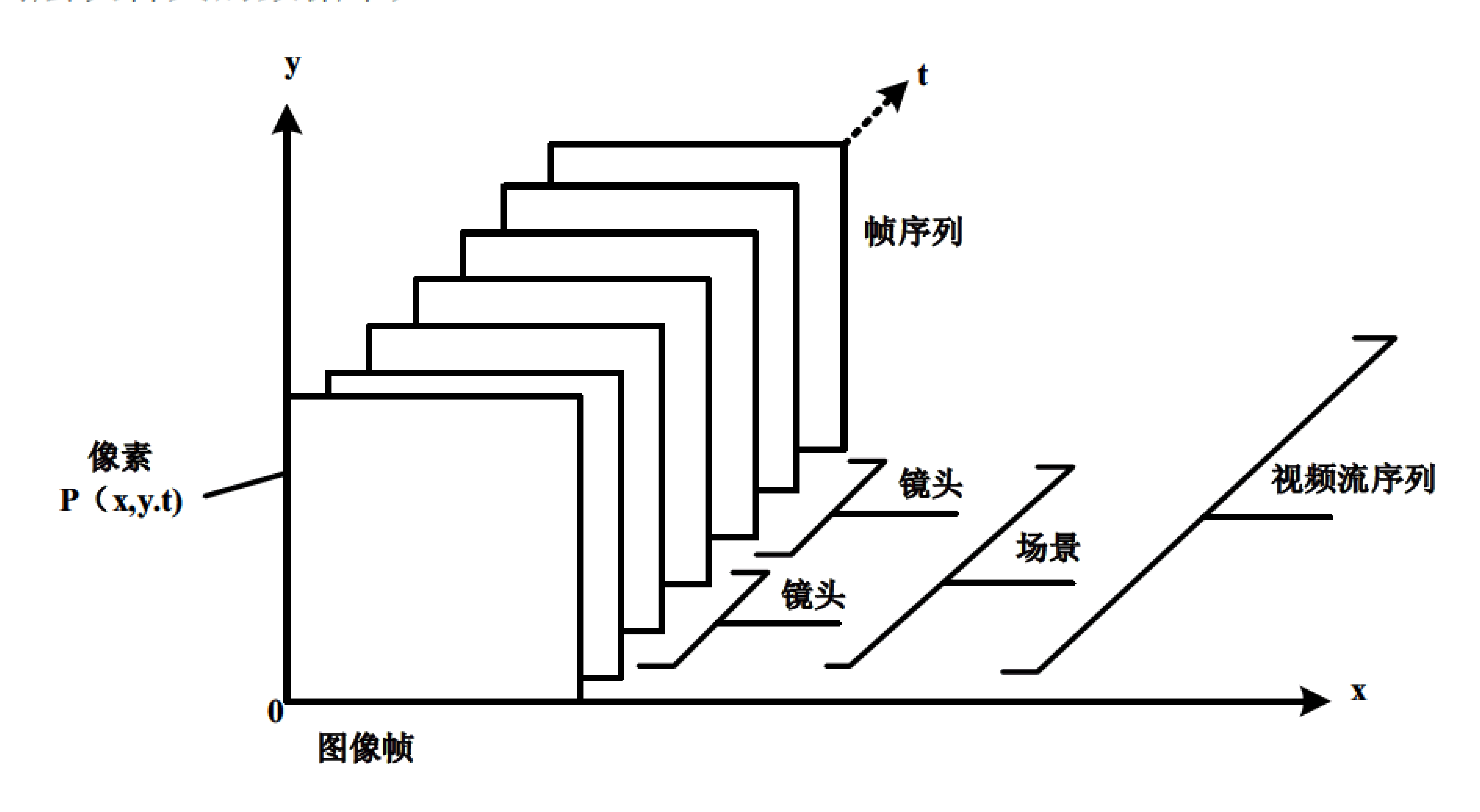


图 9.1　视频数据内容结构

1. 帧解析

帧（Frame）是视频数据的最小组成单元，是一幅静态画面。主流视频编码数据中存在 I、P、B 3 种帧类型，I 帧仅采用帧内编码模式，画面质量最好，多用于内容复杂或变化大的视频帧；P 帧和 B 帧使用帧间编码为主的编码模式，画面质量相对较低，多用于变化不大的视频帧编码。帧解析在于将原始视频数据划分为不同类型的帧序列。

算法 9.1　帧序列解析

输入：编码视频序列

过程：1. 读取视频序列头，确定视频序列的编码标准；

2. 根据编码标准，遍历所有帧的帧头数据，根据帧头中的帧类型标识位，确定该帧的具体类型。

输出：特定种类的帧序列。

---

2. 镜头解析

镜头（Shot）指一个摄像机从打开到关闭的过程中记录下来的一组连续图像帧，由

镜头边界界定，镜头解析的目的在于定位镜头切换的位置。镜头切换方式可分为两种：切变和渐变。切变指镜头和镜头之间没有任何过渡，常用检测算法主要有像素对比较法、模板比较法、颜色直方图比较法等；渐变指一个镜头以变换、溶入等方式缓慢变化到另一个镜头，常用检测算法主要有差值直方图法等。

算法 9.2　基于像素对比较的镜头解析

输入：视频帧序列

过程：1. 读取当前帧视频画面 $f_c(i,j)$，将其转换为 8 位灰度图像 $g_c(i,j)$；

2. 计算当前帧灰度图像 $g_c(i,j)$ 与前一帧灰度图像 $g_p(i,j)$ 的距离 dg_c；

$$dg_c = 1 - \frac{\sum_{i,j} \left| g_c(i,j) - g_p(i,j) \right|}{i \times j \times 255}$$

3. 如果 dg_c 大于预设阈值，则当前帧为切变镜头边界。

输出：切变镜头边界。

算法 9.3　基于灰度直方图比较的镜头解析

输入：视频帧序列

过程：1. 读取当前帧视频画面 f_c，将其转换为 8 位灰度图像 g_c；

2. 根据灰度图像生成 256 级灰度直方图 $h_c(i), i = 0,1,\cdots,255$；

3. 计算当前帧灰度直方图 $h_c(i)$ 与前一帧灰度直方图 $h_p(i)$ 的距离 dh_c，式中 N 为像素点总数；

$$dh_c = 1 - \frac{\sum_i \min\left(h_c(i), h_p(i)\right)}{N}, i = 0,1,\cdots,255$$

4. 如果 dh_c 大于预设阈值，则当前帧为切变镜头边界。

输出：切变镜头边界。

算法 9.4　基于差值直方图的镜头解析

输入：视频帧序列

过程：1. 读取当前帧视频画面 f_i，将其转换为 8 位灰度图像 g_c；

2. 根据灰度图像生成 256 级灰度直方图 $h_i(j), j=0,1,\cdots,255$；
3. 计算当前帧灰度直方图 $h_i(j)$与前一帧灰度直方图 $h_{i-1}(j)$ 的距离 dh_i，式中 N 为像素点总数；

$$dh_i = 1 - \frac{\sum_j \min\left(h_i(j), h_{i-1}(j)\right)}{N}, j = 0,1,\cdots,255$$

4. 如果 dh_i 大于预设阈值 TH_l，则当前帧为候选渐变镜头起始边界，记为 F_s；
5. 计算 F_s 与其后各帧灰度直方图的距离 ddh_{i+n}，以及各帧对应的 dh_{i+n}；

$$ddh_{i+n} = 1 - \frac{\sum_j \min\left(h_i(j), h_{i+n}(j)\right)}{N}, j = \{0,1,\cdots,255\}, n = \{1,2,\cdots\}$$

如果对于某一个 n，有如下关系成立，则 f_{i+n} 帧为候选渐变镜头结束边界，记为 F_e，其中 TH_h 为预设阈值，且有 $TH_h > TH_l$。

$$\begin{cases} ddh_{i+n} > TH_h \\ dh_{i+n} < TH_l \end{cases}$$

在得到 F_e 之前，如果对于某个 n，有如下关系成立，则清除候选渐变镜头起始边界 F_s，返回第 1 步；

$$\begin{cases} ddh_{i+n} < TH_h \\ dh_{i+n} > TH_l \end{cases}$$

6. 从 F_s 和 F_e 之间的所有帧（包括 F_s 和 F_e）中任选一帧作为渐变镜头边界。

输出：渐变镜头边界。

3. 关键帧解析

一个镜头不论长短往往带有大量冗余信息，整个视频序列表示和处理都不方便，因此需要从视频序列中提取出具有代表性的多帧，表示整个视频序列，即关键帧（Key Frame）。

最简单的关键帧选取方法是从镜头中任选一帧，如果对关键帧的提取质量要求较高，则可以采用帧平均法、直方图平均法、逐帧对比法和光流方法等。

4. 场景解析

场景（Scene）指视频中的独立故事单元，是一个高层概念。场景解析通常称为故事单元分割，对于已分割出的镜头，依据视频中的文本、声音等信息进行聚类，聚类后合并内容相近的连续镜头，得到一个单元组，称为场景信息，它可以为视频内容分析提供基础。

基本的场景解析算法步骤如下。

步骤 01 对视频进行镜头检测。
步骤 02 依据环境距离对镜头进行聚类。
步骤 03 将其中有镜头采用淡入淡出衔接方式的场景分为两个场景。
步骤 04 将场景之间的“缝隙”作为一个新的场景。

9.3 特征提取和表示

9.3.1 颜色特征提取

在视频分析中，颜色特征是应用最广泛的视觉特征，它计算简单，同时对图像本身的尺寸、方向、视角的依赖性较小。常用颜色特征包括颜色直方图、累积直方图、加权直方图和颜色矩等。

1. 颜色直方图

颜色直方图是对一幅图像中所有像素的颜色取值所作的统计，描述不同色彩在整幅图像中所占比例，不关心每种色彩所处的空间位置，可描述不需要考虑特定物体空间位置的图像内容。

如表 9.1 所示，依据不同的颜色空间，可以得到不同的颜色直方图。

表 9.1　常用的颜色空间和对应直方图取值范围

	RGB 颜色空间			HSI 颜色空间		
分量名称	R Red 红色	G Green 绿色	B Blue 蓝色	H Hue 色调	S Saturation 饱和度	I Intensity 亮度
直方图取值范围	0~255	0~255	0~255	0~359	0~100	0~255

2. 累积直方图

对于标准直方图，如果原始图像不能遍历所有可能的颜色取值，直方图中会存在较多的零值，会影响衡量直方图距离的相交运算，可考虑使用累积直方图。

在累积直方图中，每个颜色分量对应的值是所有小于等于该颜色分量的像素数所占比例，可极大减少零值出现的概率。其中 h_i 是标准颜色直方图中第 i 个颜色分量对应的值，h_i'是累积直方图中第 i 个颜色分量对应的值。

$$h_i' = \sum_{j=0}^{i} h_i$$

3. 加权直方图

人眼对于颜色空间中各个分量的感受程度存在一定差别，在实际分析时，可以为每种颜色分量附加不同的加权系数，以起到突出特定分量的作用。如对于 HIS，人眼对 H 分量最为敏感，加权系数可设为 0.7；S 分量次之，加权系数可设为 0.2；I 分量相对最不敏感，加权系数可设为 0.1。在采用加权直方图衡量两幅图像 P、Q 之间的差异时，H 分量的作用就会明显提高。

加权前：$D(P,Q) = d_H(P,Q) + d_S(P,Q) + d_I(P,Q)$

加权后：$D(P,Q) = 0.7 \times d_H(P,Q) + 0.2 \times d_S(P,Q) + 0.1 \times d_I(P,Q)$

其中 d 表示直方图间的距离。

4. 颜色矩

颜色矩是对图像颜色特征的近似，能够有效地表征图像的颜色分布，计算时无须对颜色进行量化处理，同时要能降低颜色特征的维数。在颜色矩中，颜色分布信息主要集中在低阶矩中，一阶矩μ描述平均颜色，二阶矩σ描述颜色方差，三阶矩 s 描述颜色的偏移性。其中，h_{ij} 表示第 i 颜色通道中灰度为 j 的像素出现的概率，n 表示灰度级数。

$$\mu_i = \frac{1}{n}\sum_{j=1}^{n} h_{ij}$$

$$\sigma_i = \left(\frac{1}{n}\sum_{j=1}^{n}\left(h_{ij} - \mu_i\right)^2\right)^{1/2}$$

$$s_i = \left(\frac{1}{n}\sum_{j=1}^{n}\left(h_{ij} - \mu_i\right)^3\right)^{1/3}$$

9.3.2　纹理特征提取

纹理特征包含物体表面结构排列的重要信息，通常局部呈现不规则性，整体上呈现有规律的特性。灰度共生矩阵和 Gabor 滤波器是纹理特征提取的常用手段。

1. 灰度共生矩阵

灰度共生矩阵称为空间灰度依赖矩阵（Spatial Grey Level Dependence Matrix，SGLDM），通过统计满足特定位移关系和特定灰度值的像素点对来构造矩阵，描述视频图像的纹理特征。

设 $f(x,y)$是一幅 M×N 的二维视频图像，灰度级别为 N_g，则灰度共生矩阵 $P(i,j)$为：

$$P(i,j)=\#\{(x_1,y_1),(x_2,y_2)\in M\times N \mid f(x_1,y_1)=i, f(x_2,y_2)=j\}$$

其中，#()为集合中元素的个数。

如果考虑(x_1,y_1)、(x_2,y_2)的间距 d、两点连线与坐标横轴的夹角θ，则灰度共生矩阵可扩充为 $P(i,j,d,\theta)$。

在使用灰度共生矩阵表述纹理特征时，常用的统计函数如下。

❑ 能量（Energy）

能量反映视频图像的灰度分布均匀程度和纹理粗细度，数值越大表示图像灰度分布越均匀，计算式为：

$$Energy=\sum_{i,j}P(i,j)^2$$

❑ 对比度（Contrast）

对比度反映视频图像的清晰度和纹理沟纹深浅程度，数值越大表示图像越清晰、纹理沟纹越深，计算式为：

$$Contrast=\sum_{i,j}(i-j)^2 P(i,j)$$

❑ 相关（Correlation）

相关用于度量灰度共生矩阵元素在行或列方向上的相似程度，反映图像中局部灰度相关性，计算式为：

$$Correlation = \frac{\sum_{i,j} i \times j \times P(i,j) - \mu_x \times \mu_y}{\sigma_x \times \sigma_y}$$

$$\mu_x = E\left(\sum_k P_x(i,k)\right), \mu_y = E\left(\sum_k P_y(i,k)\right)$$

$$\sigma_x = D\left(\sum_k P_x(i,k)\right), \sigma_y = D\left(\sum_k P_y(i,k)\right)$$

- 熵（Entropy）

熵用于度量视频图像具有的信息量，反映图像纹理的非均匀程度或复杂程度，计算式为：

$$Entropy = -\sum_{i,j} P(i,j) \times \log\left[P(i,j)\right]$$

- 逆差矩（IDM）

逆差矩用于度量视频图像纹理局部变化程度，反映图像纹理的同质性，数值越大表明图像区域变化越小，计算式为：

$$IDM = \sum_{i,j} \frac{1}{1+(i-j)^2} \times P(i,j)$$

2. Gabor 滤波器

Gabor 滤波器是在 Fourier 变换的基础上增加一个 Gaussian 窗口函数，可以通过不同尺度和方向滤波器的设计来反映图像局部像素分布特征，对图像纹理有非常强的描述能力。通常的做法是设计合适的 Gabor 滤波器去过滤图像，对过滤后的图像提取能量统计特征作为纹理特征，一种有效的 Gabor 纹理表示 $g(x,y)$及对应的 Fourier 变换 $G(u,v)$为：

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right]$$

$$G(u,v) = \exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right\}$$

通过对 $g(x,y)$进行膨胀（比例因子 z ）和旋转（角度θ），可得到一组 Gabor 滤波器 $g_{z\theta}(x,y)$用于纹理特征提取：

$$g_{z\theta}(x,y)=z\times g(x',y')$$
$$x'=z\times(x\cos\theta+y\sin\theta)$$
$$y'=z\times(-x\sin\theta+y\cos\theta)$$

9.3.3　形状特征提取

形状是物体的基本特征之一，形状描述方法主要有 Fourier 描述子、曲率描述子、Zernike 矩等。

1. Fourier 描述子

设 $\{(x_k,y_k),k=0,1,\cdots,K-1\}$ 是构成二维平面中封闭边界的点集，将其用复数形式转化为一维序列 $s(k)$。

$$s(k)=x_k+jy_k$$

对 $s(k)$做 Fourier 变换，得到边界的 Fourier 描述子 $S(u)$。

$$S(u)=\frac{1}{K}\sum_{k=0}^{K-1}s(k)\times e^{-j2\pi uk/K}$$

$S(u)$的高频分量对应轮廓的细节分量，低频分量对应轮廓的基本形状，因此可以采用少量的低频 Fourier 系数即可实现图像轮廓的重建。归一化的 Fourier 描述子具有旋转、平移和缩放不变性，并且与轮廓的起点无关。

2. 曲率描述子

使用曲线的弧长 l 为参数对闭合轮廓曲线的平面坐标 x、y 进行参数化。以任意一点为起点，顺时针跟踪轮廓，并对 l 进行归一化，将轮廓曲线表示为：

$$C=\{x(l),y(l)\},l\in[0,1]$$

曲率描述子 $k(l)$的计算式为：

$$k(l)=\frac{\dot{x}(l)\times\ddot{y}(l)-\ddot{x}(l)\times\dot{y}(l)}{\left(\dot{x}^2(l)+\dot{y}^2(l)\right)^{3/2}}$$

$$\dot{x}=\frac{dx}{dl},\ \dot{y}=\frac{dy}{dl}$$

$$\ddot{x}=\frac{d^2x}{dl^2},\ \ddot{y}=\frac{d^2y}{dl^2}$$

3. Zernike 矩

Zernike 矩是一种正交复数矩，所利用的正交多项式集是一个在单位圆内的完备正交集，其定义为：

$$Z_{mn}=\frac{m+1}{\pi}\iint_{x^2+y^2\le 1}\left[V_{mn}(x,y)\right]^* f(x,y)dxdy$$

式中 $f(x,y)$为原始图像，$V_{mn}(x,y)$ 为 Zernike 多项式，“*”代表复共轭，$m=\{0,1,\cdots,\infty\}$，n 为整数，且有$(m-|n|)$为非负偶数。

Zemike 矩的基 $V_{mn}(x,y)$ 是正交径向多项式，可以保证所提取特征的相关性小、冗余性小、抗噪声能力强，且具有平移不变性。一幅图像的形状特征可以用一组 Zemike 矩特征向量很好地表示，其中低阶矩描述整体形状，高阶矩描述目标细节。

9.3.4 运动特征提取

运动特征指视频中随时间变化的特征，主要由两部分组成，一是反映摄像机运动的背景运动特征，二是目标运动的前景运动特征，这些特征对视频内容描述和理解非常重要，是视频数据独有的特征。

1. 背景运动特征

摄像机的运动主要有 7 种：Panning（左右转动）、Tilting（上下转动）、Zooming（焦距变化）、Tracking（水平追踪）、Booming（垂直追踪）、Dollying（前后追踪）、Rolling（绕光轴旋转）。如图 9.2 所示，提取全局运动特征常用基于参数模型的全局运动估计方法。

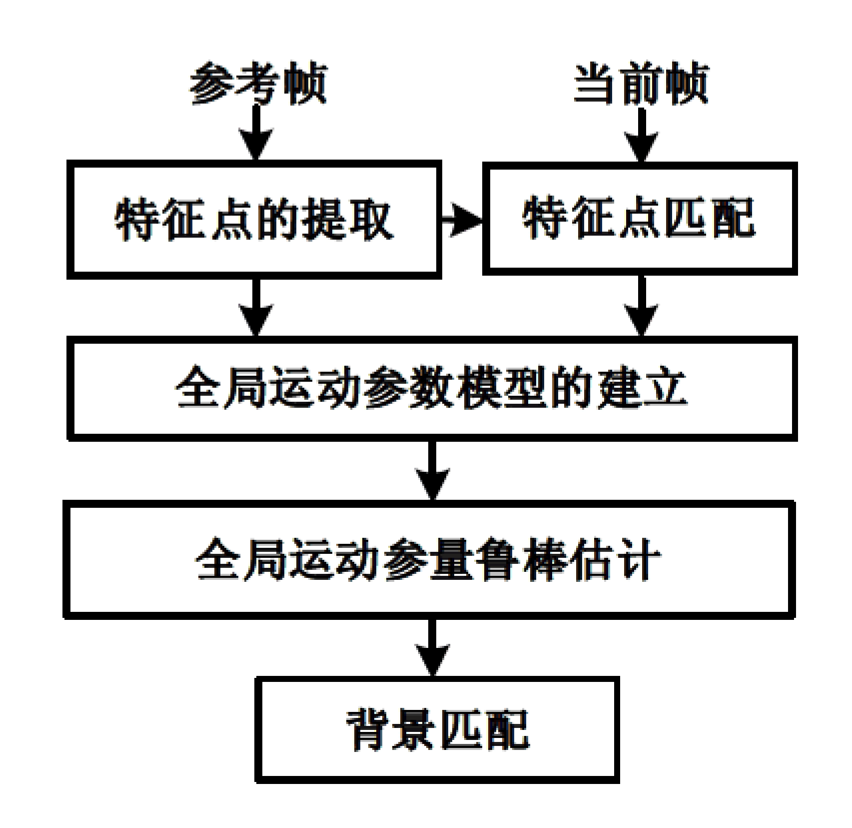


图 9.2　基于参数模型的全局运动估计

常用特征点提取方法有 SUSAN、Harris 和 SIFT 等算子，其中 Harris 算子计算简单、稳定，应用广泛，计算方法如下。

步骤 01　图像求导，其中 I_x、I_y 分别对应像素点在 x、y 方向的倒数。

$$M=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\otimes\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix}$$

$$\xRightarrow{\text{对角化}} \mathrm{R}^{-1}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}R$$

步骤 02　计算 M 的行列式 det 和秩 Tr。

$$\det(M)=\lambda_1+\lambda_2$$

$$Tr(M)=\lambda_1\times\lambda_2$$

步骤 03　计算 Harris 算子，其中 k 为默认常数（常为 0.02~0.04）。

$$\mathrm{R}=\det(M)-k\times Tr^2(M)$$

步骤 04　当 R 大于预设阈值且为局部极值时，该点为所求特征点。

下面是特征点匹配常用的模板法。

算法 9.5　基于模板法的特征点匹配

输入：属于不同帧的两个特征点 $p(x_i,y_i)$、$q(x_j,y_j)$

过程：1. 在两帧中以 $p(x_i,y_i)$、$q(x_j,y_j)$为中心，R 为半径划定待匹配区域 P_i、Q_j；

2. 计算 P_i 和 Q_j 的相似度 SAD_{ij}；

$$SADi_j = \sum \left| P_i(x,y) - Q_j(x,y) \right|$$

3. 如果 SAD_{ij} 小于预设阈值，则 $p(x_i,y_i)$、$q(x_j,y_j)$构成匹配点对。

输出：$p(x_i,y_i)$、$q(x_j,y_j)$的匹配结果。

摄像机的运动使视频图像中像素点的坐标从 $k-1$ 帧的(x_{k-1},y_{k-1})处移动到 k 帧的(x_k,y_k)处，坐标变换量满足一定运动变换模型，即：

$$(x_{k-1}, y_{k-1}) = f(x_k, y_k)$$

常见变换模型如下。

- 二参数运动模型

可表征平移运动，(c,d)为沿坐标轴的偏移量。

$$\begin{cases} x_{k-1} = x_k + c \\ y_{k-1} = y_k + d \end{cases}$$

- 四参数仿射运动模型

可表征平移、旋转、伸缩运动，γ为缩放参数，θ为旋转角度。

$$\begin{pmatrix} x_{k-1} \\ y_{k-1} \end{pmatrix} = \gamma \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}$$
$$= \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}$$

在求得匹配点对并建立变换模型之后，可通过最小二乘法估计变换模型中的参数，得到全局运动特征。

2. 前景运动特征

前景运动特征的提取分为 3 步：运动目标分割、运动目标跟踪和运动特征提取。

- 运动目标分割

将运动前景与背景分离，具体步骤如下。

步骤01 对输入视频数据进行全局运动检测，对存在全局运动的数据进行运动补偿，消除摄像机运动的影响。

步骤02 采用时间差分法、背景减除法或光流法提取前景运动目标。

- 运动目标跟踪

在一段视频序列中将隶属于同一运动目标的区域分割出来，用于后续运动特征提

取。相关算法参见 7.2.1 小节。

❑ 运动特征提取

在得到运动目标序列之后，可提取目标的运动特征。

算法 9.6　瞬时全局运动速度与方向提取

输入：相邻两帧视频画面中隶属于同一运动目标的区域 A_i、A_{i-1}。

过程：1. 提取 A_i、A_{i-1} 的重心 cg_i、cg_{i-1}；

2. 连接 cg_{i-1} 与 cg_i，得到当前帧目标整体运动矢量 M_{Ai}；

3. M_{Ai} 的指向即为目标瞬时运动方向，

M_{Ai} 的长度和帧间隔之比为目标瞬时运动速度。

输出：当前帧目标瞬时运动方向和速度。

算法 9.7　瞬时全局运动速度变化量和方向变化量提取

输入：相邻两帧视频画面中隶属于同一运动目标的整体运动矢量 M_{Ai}、M_{Ai-1}。

过程：1. 计算 M_{Ai} 和 M_{Ai-1} 的差值；

$$DM_{Ai} = M_{Ai} - M_{Ai-1}$$

2. DM_{Ai} 的指向为运动方向的变化量，$\mathrm{D}M_{Ai}$ 的幅值和帧间隔之比为运动加速度。

输出：当前帧目标瞬时运动方向变化和加速度。

算法 9.8　运动轨迹提取

输入：视频序列中隶属于同一运动目标的区域$\{A_i\}$。

过程：1.对每一个 A_i，提取其质心 cg_i；

2. 将所有 cg_i 按时间顺序相连，得到目标运动轨迹。

输出：目标运动轨迹。

9.3.5　音频特征提取

与视频内容同步的音频特征能够表征视频内容的重要程度，如在视频监视系统中，当呼救、大声喊叫、碰撞声、枪声等异常声音出现时，意味着此时的视频内容中可能包

含值得注意的异常行为。

下面介绍常用的音频特征。

1. 梅尔频谱系数（MFCC）

鉴于人的听觉特性，提取音频特征时，MFCC（Mel-Frequency Cepstrum Coefficients）利用 Mel 频率刻度，对声音频率进行变换。具体地，依据人耳感受声音时声音高低和频率间的非线性关系，实现声音信号频谱到基于 Mel 频率点的非线性频谱的转换，最终再转换到倒谱域上。

在 MFCC 转换过程中，考虑到频率轴上的 FFT 变换谱线等间隔分布，FFT 变换谱线经常使用一组滤波器组进行滤波，该滤波器组依据人耳听觉的临界频带分布进行设计，中心频率尽管在频率轴上非均匀分布，但在 Mel 频率轴上却是等间隔分布的，其非线性特性与人耳听觉相似。

2. 短时过零率

过零率是信号频谱特性的反映。离散时间信号的波形与零电平的横轴相交时，称为“过零”，此时信号的两个相邻采样点的符号相反。

平均过零率可以通过计算单位时间内采样点符号的改变次数得到。短时平均过零率定义为：

$$Z_n = \sum_{m=-\infty}^{\infty} \left| \operatorname{sgn}[x[m] - \operatorname{sgn}[x(m-1)] \right| w(n-m)$$

其中，sgn[]为符号函数，$w(n-m)$为窗口函数。

$$\operatorname{sgn}\left|x(n)\right| = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases}$$

在矩形窗条件下，可以简化为：

$$Z_n = \frac{1}{2N} \sum_{m=n-N+1}^{n} \left| \operatorname{sgn}[x(m) - \operatorname{sgn}[x(m-1)] \right|$$

3. 短时帧能量

声音信号的能量随着时间变化比较明显，声音信号的短时能量分析能够很好地描述幅度变化。短时能量的波形随着声音信号的幅度而变化，能很好地体现声音信号的时域信息。

短时帧能量等于该段语音取样值的平方和，在实际应用中可用平均幅值代替。

4. 基音周期

声音中浊音信号的周期称为基音周期，是振动频率的倒数，基音周期的估计称为基音检测，通常利用自相关函数进行基音检测。对于离散的数字声音信号序列 $x(n)$，自相关函数为：

$$R(k)=\sum_{m=-\infty}^{\infty}x(m)x(m+k)$$

其中，k 为声音信号的延迟点数。对于随机信号或周期信号序列，自相关函数定义为：

$$R(k)=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{m=-N}^{N}x(m)x(m+k)$$

如果序列 $x(n)$具有周期 N_p，则自相关函数是同周期的周期函数，即：

$$x(n)=x(n+N_p)$$

则

$$R(k)=R(k+N_p)$$

5. 带宽

带宽为取样信号的频率值范围，用于表征音频信号的类型。

9.4　典型系统

视频摘要最早可追溯到 20 世纪 90 年代中后期美国卡内基梅隆大学开发的 Informedia 工程，德国曼海姆大学的 MoCA 系统、美国 IBM 的 QBIC 系统、美国 FX Palo Alto 实验室的 Video Manga 系统、新加坡国立大学的 SWIM 系统等都是具有代表性的视频摘要系统。

北京大学、清华大学、浙江大学、中科院自动化所、国防科技大学等单位深入研究视频摘要，其中国防科技大学海量视频分析与安全预警研究中心（VAP）研发的“面向安全监视的海量视频摘要系统”获得了第七届国际发明展览会金奖。

1. Informedia: News-on-Demand

近年来互联网的多媒体数据呈现爆炸式增长，美国卡内基梅隆大学开展了Informedia工程，创建能够对文字、图像、音频、视频内容进行完全检索的数字图书馆。

Informedia工程综合自然语言理解、图像处理、语音识别和视频压缩等领域的相关研究成果，极大提高了用户使用多媒体信息的深度和广度。它将视频数据分割为逻辑片段，根据逻辑片段所包含的具体内容生成对应的索引。用户在查找视频信息时可以直接搜索索引信息，并快速跳转到所需的逻辑片段。用户通过Informedia数字图书馆进行信息检索时可以直接输入关键字（通过键盘或麦克风）或者选择系统中预设的分类条目，系统可以智能识别用户的输入请求，并选择和用户要求最相关的内容发送给用户。

News-on-Demand（新闻点播）是Informedia工程的一个具体应用，能够自动从视频、音频、文字媒体中抓取用户感兴趣的新闻内容。新闻是一种时效性很强的数据，每时每刻都有新的新闻数据产生，依靠人力管理这些数据是异常艰巨的任务。News-on-Demand借助先进的计算机技术，极大提高了对新闻数据的管理能力。

News-on-Demand的基本工作流程和相关技术（图9.3）如下。

步骤01 对音频、视频数据进行数字化和压缩编码（MPEG-X）。

步骤02 依据视频字幕或语音识别结果创建视频的时间线（HMMs、码书）。

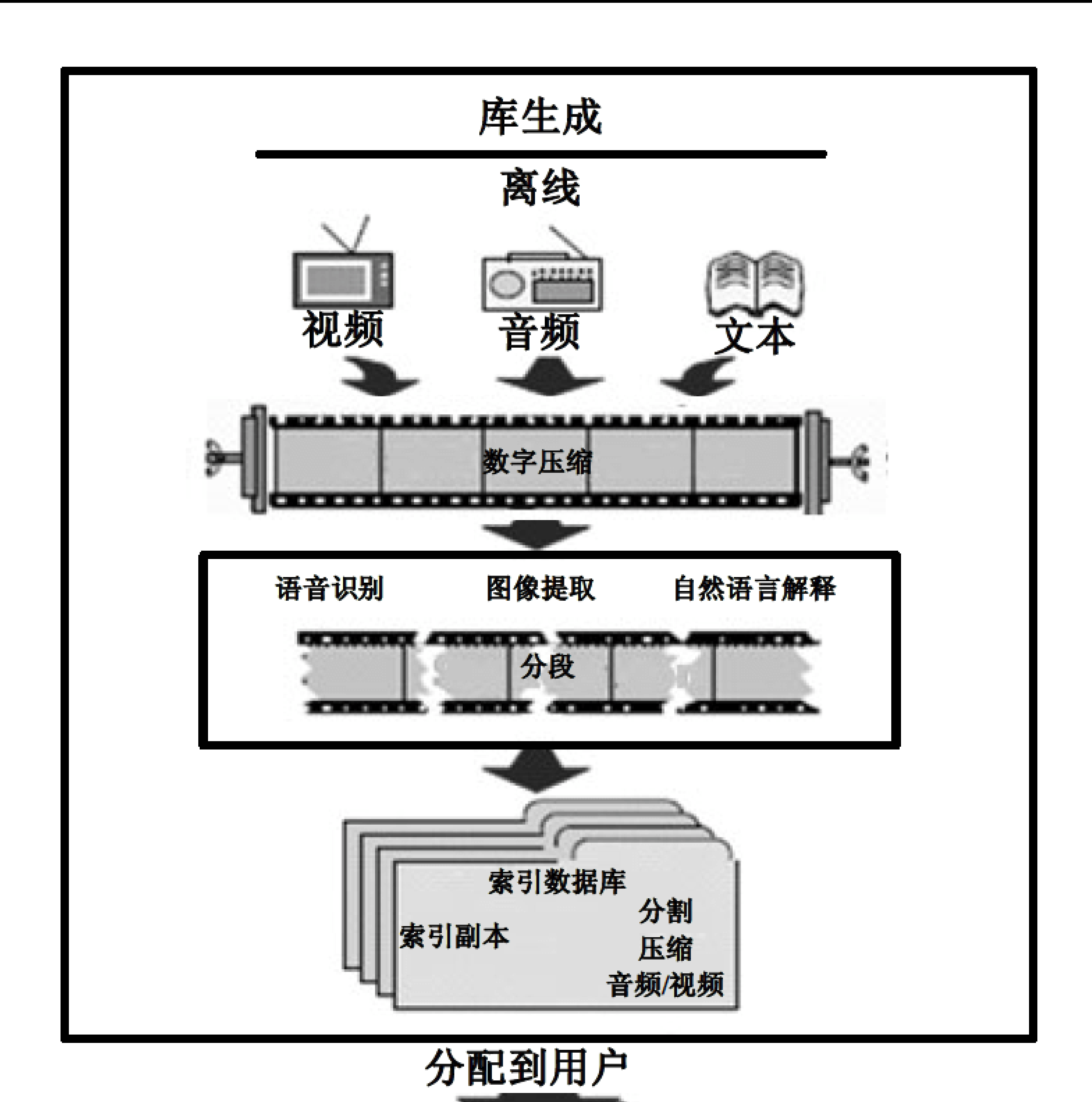
步骤03 分割故事边界（基于颜色直方图的场景分割）。

步骤04 分割场景关键帧（基于光流法的运动目标检测）。

步骤05 使用Informedia系统对视频形成索引。

2. QBIC

QBIC（Query By Image Content）是IBM开发的基于内容的海量多媒体数据检索系统。通过分析颜色、形状、纹理和骨架特征，使用户可以从海量视频、图像数据库中检索到特定信息。其架构如图9.4所示。



分配到用户

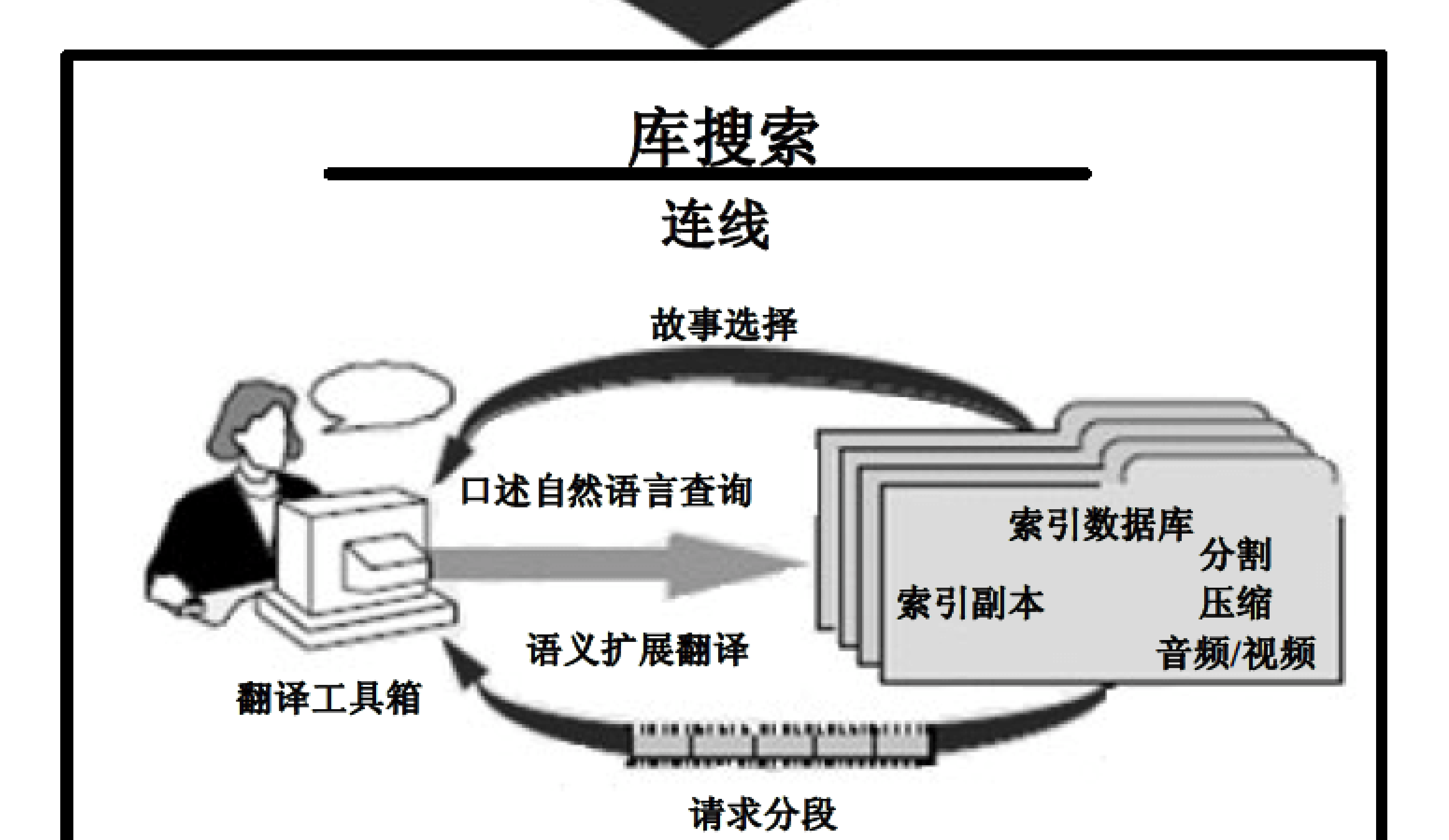


图 9.3　Informedia：News-on-Demand 架构

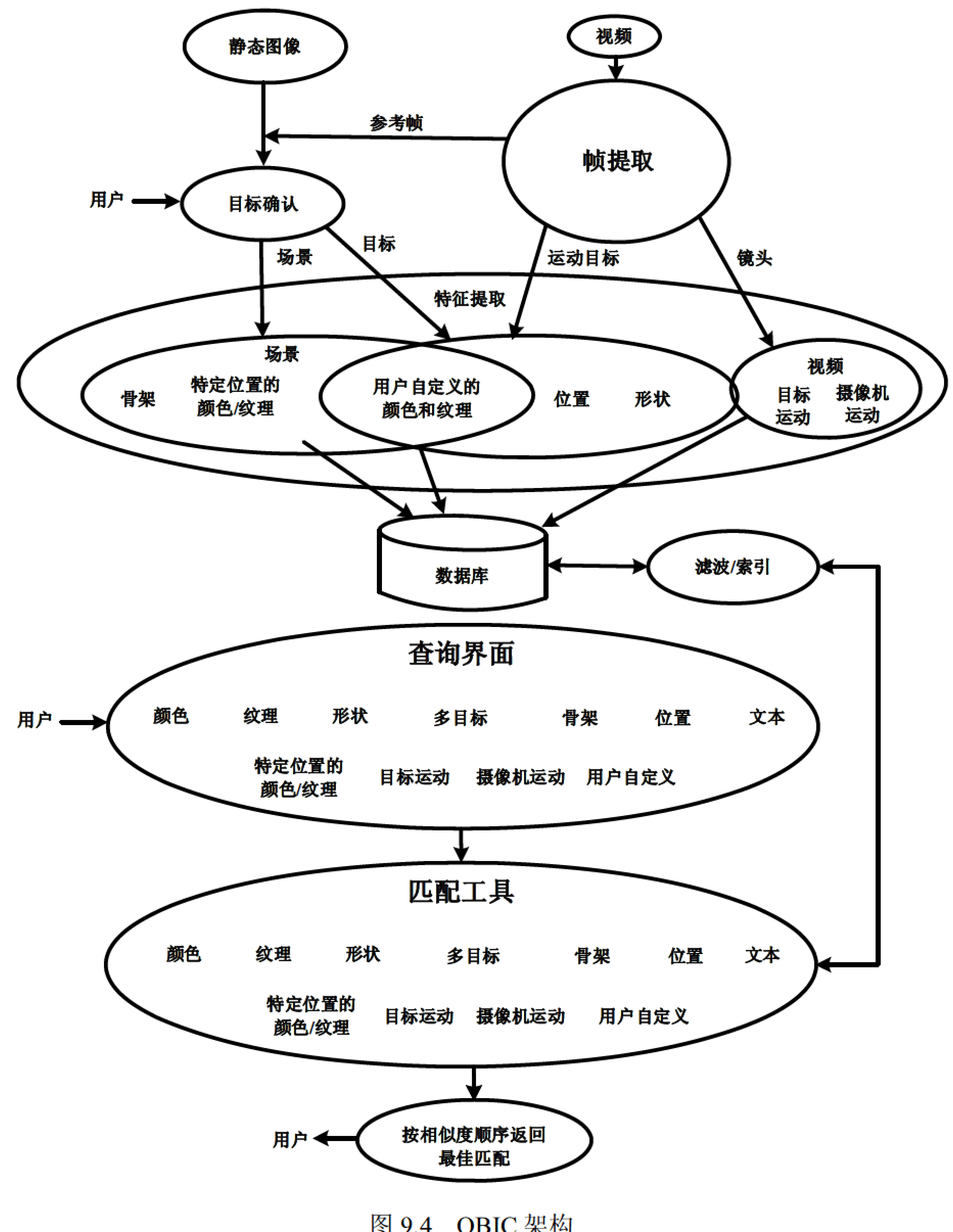


图 9.4　QBIC 架构

QBIC 系统支持如下检索方式：

- 用户给出模板图片，系统根据图片搜索；
- 用户绘制出骨架草图，系统根据骨架草图搜索；
- 用户从颜色、纹理集中选取目标特性，系统根据此特性搜索。

3. Video Manga

Video Manga 是美国 FX Palo Alto 实验室开发的视频摘要系统，基本工作流程如下。

步骤01 通过分析视频图像和对应的音频特征，对视频画面进行聚类，将原始视频分为多个故事$\{S_i\}$，每个故事由若干个片段$\{C_{ij}\}$组成。

步骤02 以故事的长度 L_{Si} 为标准，计算每个故事的权重 W_i。

$$W_i = \frac{L_{Si}}{\sum_i L_{Si}}$$

步骤03 计算每个视频片段的重要性 I_{ij}。

$$Ii_j = Li_j \log W_i$$

步骤04 依据片段的重要性选取出关键帧。

步骤05 依据故事的权重设定关键帧的尺寸，并将所有关键帧组合成漫画形式的视频摘要。

图 9.5 显示该系统的一个实例。

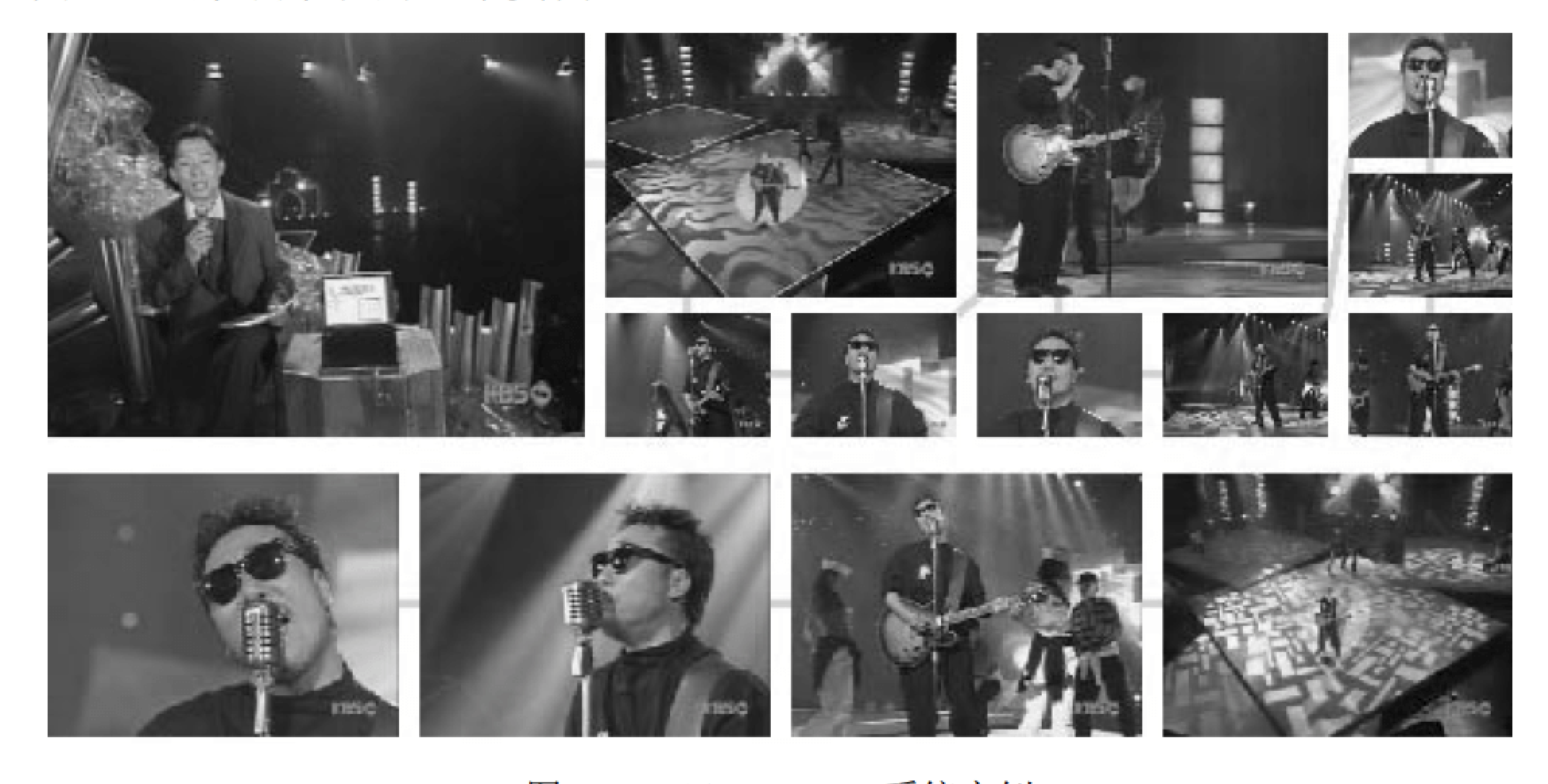


图 9.5　Video Manga 系统实例

第 10 章

海量视频管控平台

海量视频管控平台面向视觉大数据，基于海量视频模型，采用 Hadoop 等数据处理框架，通过视频分析方法，给用户提供友好的、可视的、智能的海量视频管理和操控工具。

本章以某地级市为例，详细阐述基于海量视频管控平台的视频监控与回放、视图无缝融合、大规模人脸等目标监测、异常行为检测、海量视频摘要、高清卡口车辆信息搜索等功能。

10.1 平台要求

海量视频管控平台有以下 3 点要求。

1. 先进性

采用先进的视频处理、分析与理解技术，支持高清图像大数据量的稳定传输功能，支持海量视频数据的高效解码和快速识别功能，支持高清视频的高画质可视化功能。

2. 安全性

很多海量视频数据涉及国家安全和公民隐私，平台应该具有防范计算机病毒的能力，有很强的抗干扰能力，具有授权密码、多级控制、设防级别等功能，避免遭遇恶意

攻击、非法提取数据等违法行为。

3. 兼容性

在传输协议、数据接口、SDK 控件、记录结构等多层面，支持对符合标准的模块、设备、数据库、子系统等无缝接入。对于设备类的对接，采用标准的网络协议，稳定、高效地接入到平台之中。对于数据库类的交换，提供数据分类导入、导出功能，方便海量视频数据共享。

10.2　平台架构

如图 10.1 所示，海量视频管控平台从上至下共分 4 个层次，分别是用户界面层、业务应用层、系统服务层和设备接入层。

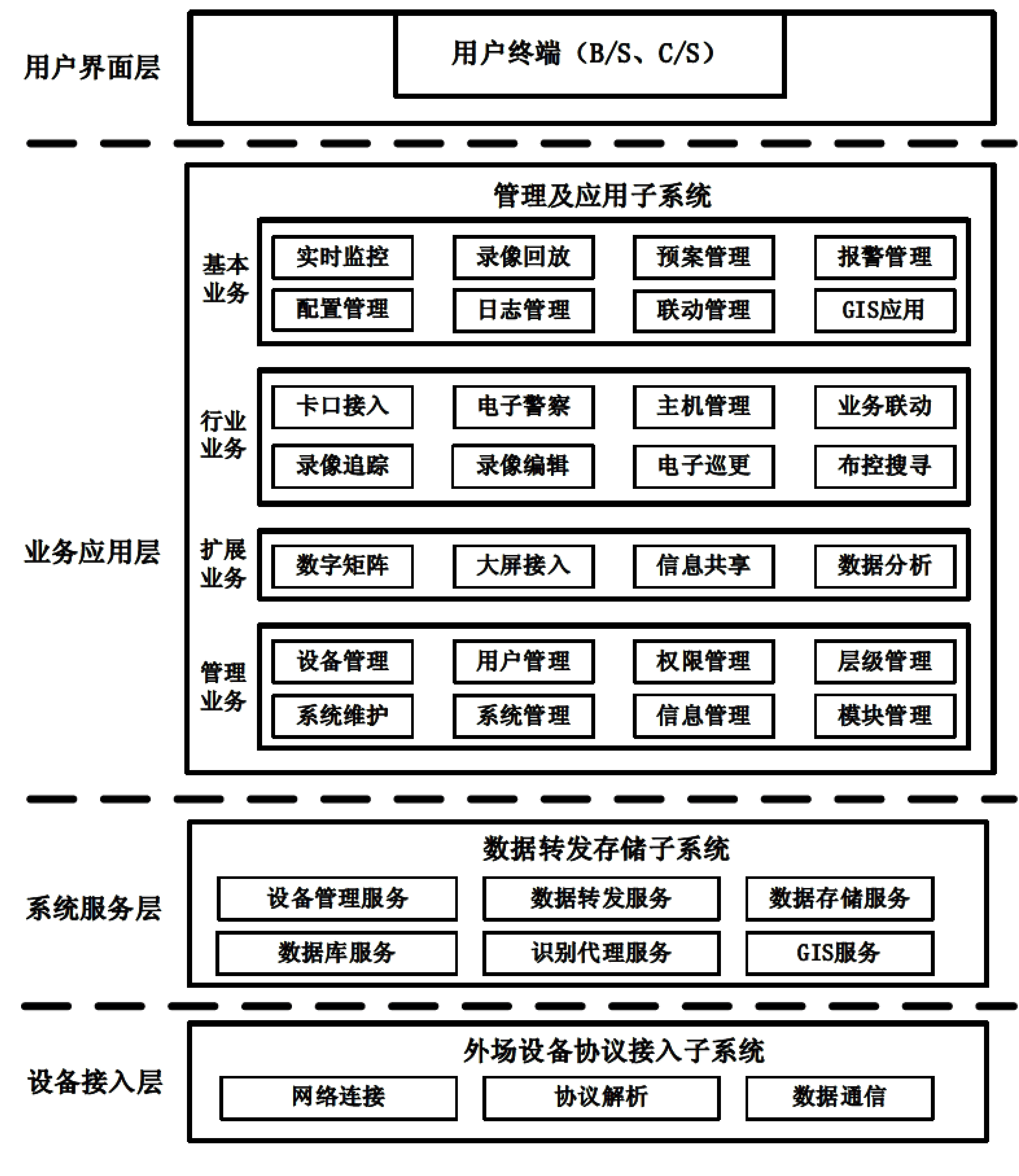


图 10.1　平台架构

1. 设备接入层

海量视频管控平台要实现数据综合、处理、流转、运行，90%以上的数据由前端采集设备收集、传输而来，平台需要与这些设备连接、交互。

该平台单独封装设备接入层，完成服务器和前端设备、客户端等之间的网络连接、协议解析、数据通信等功能。

2. 系统服务层

该平台面向庞大数据库，独立封装功能服务器，使整个服务器结构灵活，规模可扩展性强，功能之间耦合性低，使整个系统稳定、有序、高效运行。

系统服务层提供中心管理、设备管理、媒体转发、媒体存储、图像管理、目标识别等多种服务。

3. 业务应用层

业务应用层面对用户的系统客户端功能呈现，方便用户对系统设备、用户、任务的管控，实现了系统整体性、用户便捷性、运行稳定性。

业务应用层的业务包括：实时监视、录像回放、日志管理等视频监控业务；卡口管理、电子警察管理、车牌识别、布控撤防等交通管理业务；大屏接入、数字矩阵等人机交互业务；设备管理、用户管理、权限管理、录像计划等系统管理业务。

4. 用户界面层

平台采用 B/S 展现模式，便于异地浏览，只要连通网络，可以把任何计算机看做客户端，在任何时间、任何地点、任何系统中，使用浏览器直接连接服务器。

视频显示模块采用 C/S 的嵌入插件方式，该方式具有信息采集灵活、负载均衡、服务稳定的优点，增强了客户端的事务处理能力，减轻了服务器的工作负担。

10.3 平台组成

在前端识别模式中，视频图像识别工作由前端设备完成，如带识别功能的摄像机。在中心识别模式中，视频图像识别工作在管理中心完成，平台结构如图 10.2 所示。

常用的是综合识别模式，系统同时带有前端识别和中心识别模式，在遇到前端繁忙、计算能力弱等情况时，在视频图像传输到中心后，再进行二次识别。

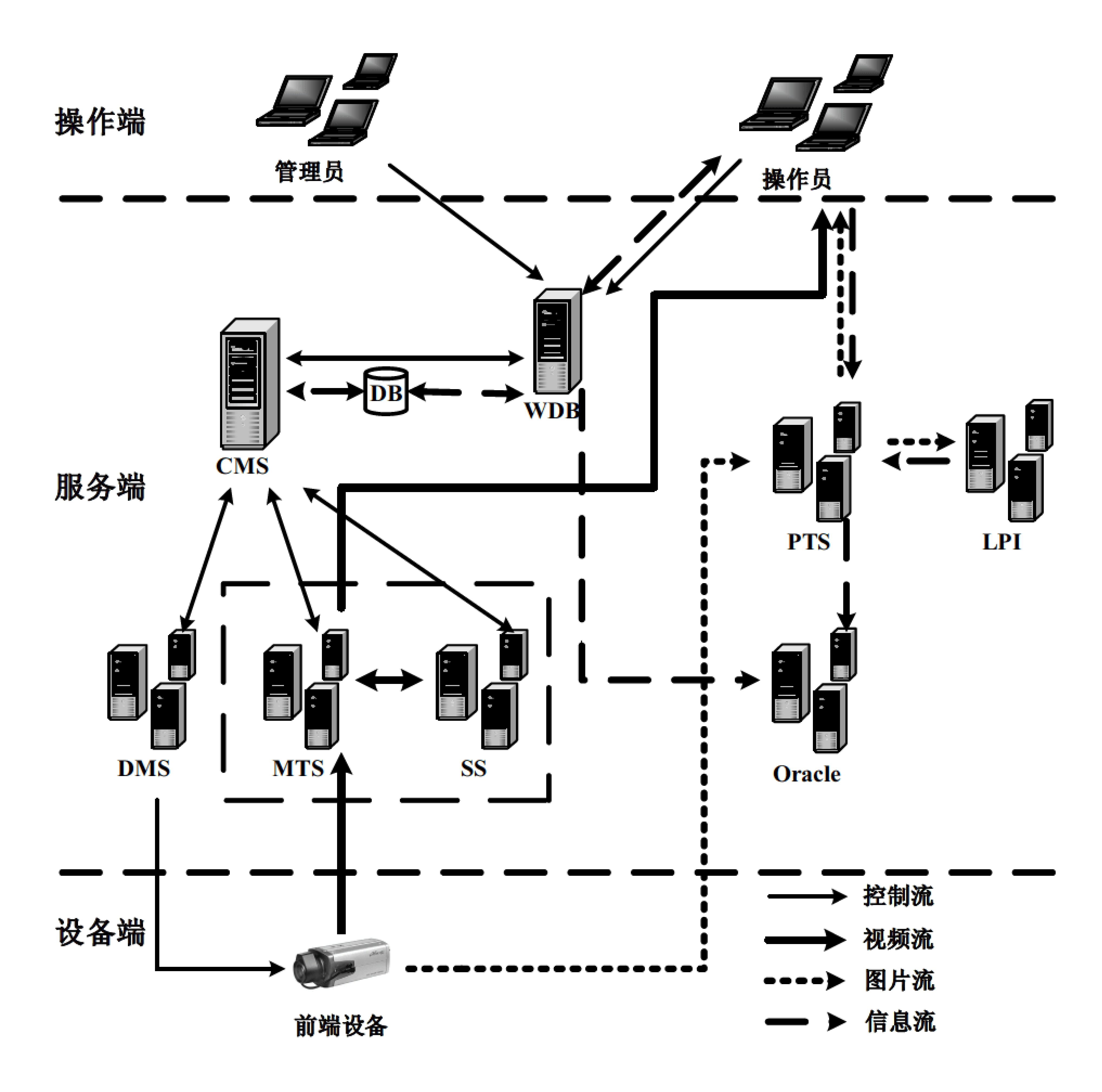


图 10.2　中心识别模式的平台结构

如图 10.3 所示，平台软件提供核心业务管理、媒体转发、音视频存储、设备管理、数据库管理、图像管理、Web 等服务。

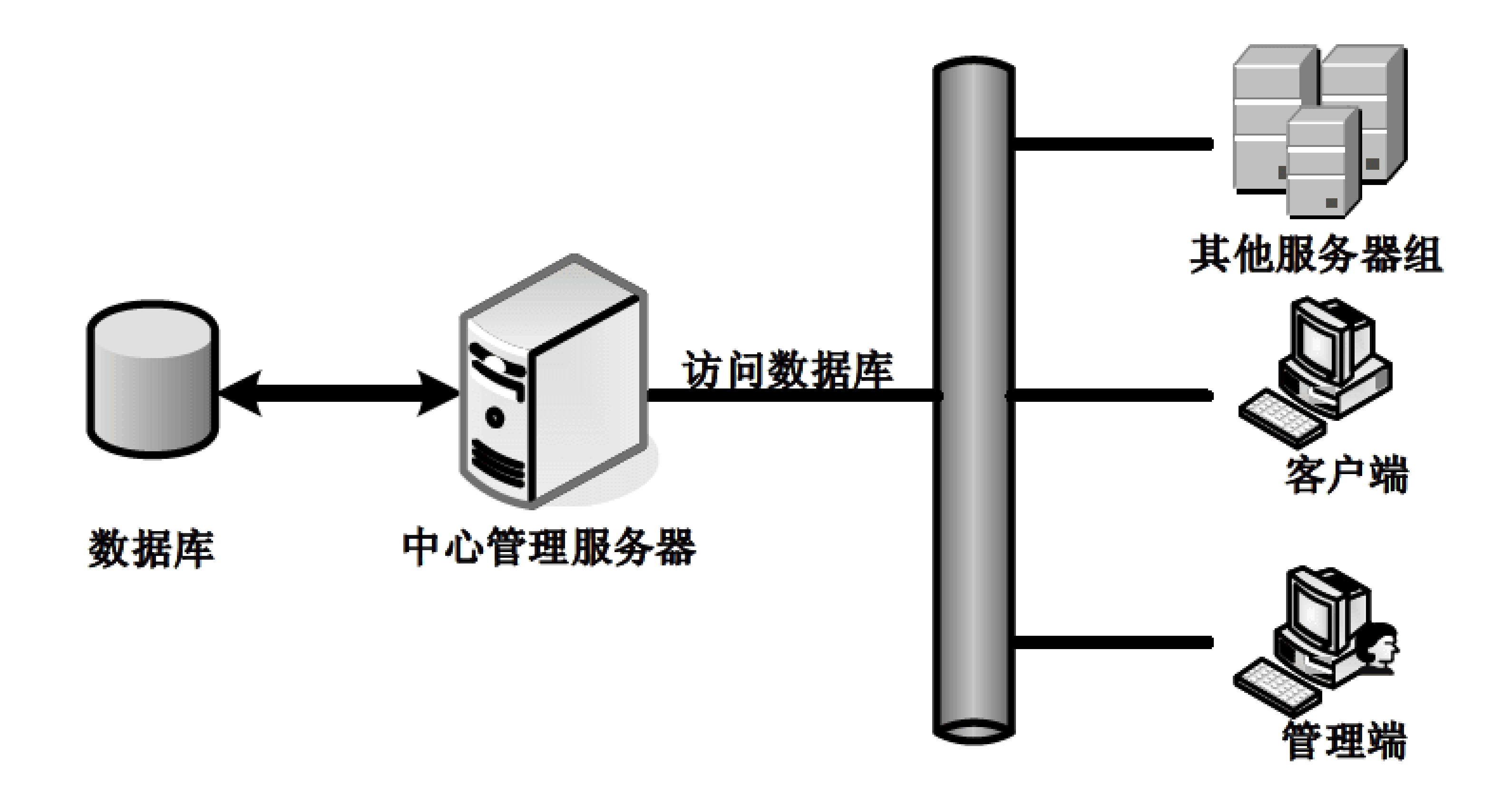


图 10.3　平台服务组成框图

1. 核心业务管理服务（CMS）

CMS（Centre Manage Server）负责业务监控、权限控制、系统容错、负载均衡、动态集群等工作。针对不同的业务逻辑的需要，CMS 给 DMS（Device Manage Server）、MTS（Media Transmit Server）、存储服务器等发送不同的命令，执行相应的处理。

2. 媒体转发服务（MTS）

MTS 的任务是从前端设备处获取音视频数据，并按照标准流媒体协议，将数据转发给存储服务器和客户端，支持一对一、一对多和多对多三种转发模式，支持视频流相关的统计信息。

3. 音视频存储服务（MSS）

MSS（Media Store Server）采用虚拟存储管理技术，支持 DAS、NAS、IP-SAN 等存储设备；支持标准的 NFS、SAMBA、ISCSI 等文件协议；支持 PB 级海量音视频数据存储、快速检索。

4. 设备管理服务（DMS）

DMS 负责设备的管理工作，向设备发送命令（如查询、配置、操作）、收集设备的网管信息和报警信息、实施报警联动策略。

5. 图像管理服务（PMS）

PTS（Picture Manage Server）负责将图像保存在数据库服务器上，供客户端实时监控和查询。它支持标准的 SAMBA、NFS、ISCSI 等文件协议，支持 NAS、DAS、FC-SAN、IP-SAN 存储方案。

6. Web 服务（WBS）

WBS（Web Server）以 Web 形式向客户提供 Web 访问功能，方便与其他子系统接口。采用 B/S 架构，通过 IE 进行访问，Web 端集成客户端的基本功能及部分管理端功能，实现实播、回放、配置等结果。

7. 解码上墙服务（VMS）

VMS（Video Manage Server）实现视频解码、显示，可连接至 DLP、LED 墙等。

8. 车牌识别服务（LPI）

车牌识别服务（License Plate Identification）将抓拍图像及识别信息传给图像管理服务器，如果遇到未识别图像，则传输给车牌识别服务器，并将识别信息记录在数据库中。

10.4　平台服务器

如图 10.4 所示，平台服务器及存储设备连接示意图如下。

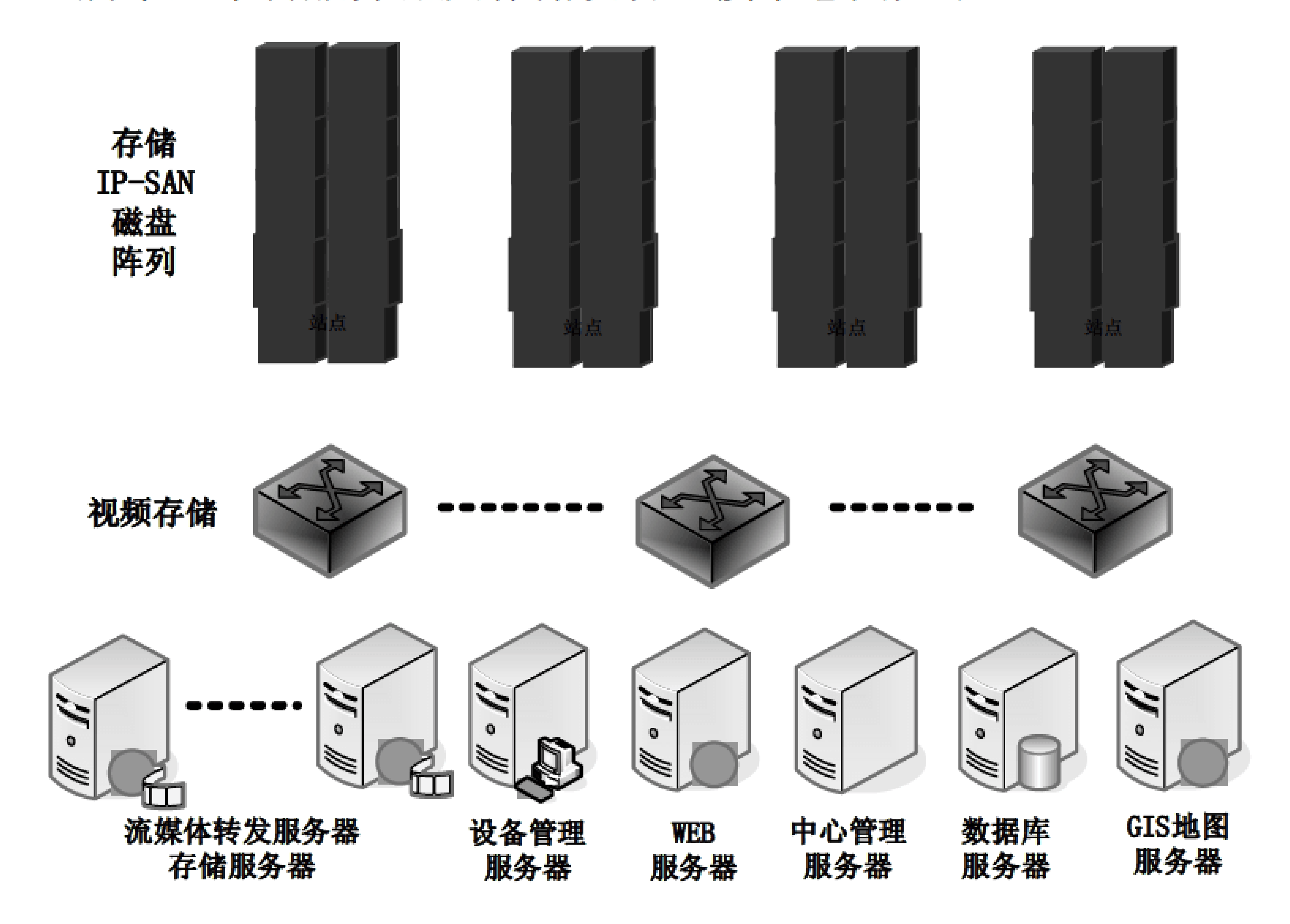


图 10.4　平台服务器及存储设备

以某地级市为例，要求监控视频数据都保存在监控中心，保存时间为 30 天，视频数据都从监控中心调用，方便资源共享。

每台标清摄像机（D1）的码流在 0.2MB 左右，按照 0.2MB 码流计算每路存储所需

空间如下。

- 每路每天的数据量为 0.2×60×60×24≈16.9GB
- 存储 30 天的存储容量为 16.9×30＝507GB

每台高清摄像机（1080P/720P）的码流在 1MB 左右，按照 1MB 码流计算每路存储所需空间如下。

- 每路每天的数据量为 1×60×60×24≈84.4GB
- 存储 30 天的存储容量为 84.4×30≈2.53TB

若该地级市共有 10,000 台标清摄像机和 10,000 台高清摄像机，则存储 30 天的存储容量为 30,370TB，约 30PB。考虑到监控中心对于存储数据的大流量、高反应速度要求，存储系统可使用 IP-SAN 架构的高性能存储设备。

10.5　平台功能

10.5.1　视频监控与回放

1. 视频实时监控

如图 10.5 所示，前端主机直联、服务器转发的视频数据，均可在客户端实时播放；支持视频双码流传输；可按照指定设备、指定通道远程监听任意某路音频信号，同时记录多个监听通道的音频信号。

支持视频实时浏览和切换控制，支持多画面组合模式（如单画面、九画面、三十二画面）监控，支持图像抓拍和视频录像。可按照指定场所、通道进行单路图像、报警联动图像的实时点播及轮循切换显示。

设备树分级显示所有设备，采用不同图标显示设备的不同状态，实时刷新设备状态，快速发现设备故障。

具有视频切换功能，可在指定的显示器上实时显示指定摄像机的监控视频。

具有云台镜头控制功能，控制云台转动、镜头光圈和变倍聚焦、预置点操作。

可对可疑目标进行三维智能定位，将其定位在屏幕中心，并对目标区域进行适当缩放，快速锁定可疑目标，及时发现可疑现场并保存视频证据。

可对视频图像进行放大、缩小操作，调整图像亮度、对比度和色度等属性，将视频显示效果调整到最佳状态。



图 10.5　平台的视频实时监控

2. 视频录像回放

如图 10.6 所示，视频录像的快速检索、流畅播放是平台的重要功能，便于事发后有据可查。

可回放设备存储录像，或者平台存储录像，可以支持多路不同的录像同时回放。

支持录像下载到客户端，可按照时间或者文件下载。按照时间下载时，可以精确到秒；按照文件下载时，可以采用打包方式，便于批量下载。

采用不同颜色标注不同类型的录像，突出重点视频。支持移动侦测、外部报警、视频遮挡、视频丢失等自动检测功能。

可按百分比或时间显示录像进度条，可跨文件连续播放，支持停止、暂停/播放、逐帧播放、快放/慢放等功能，支持音量大小调节。



图 10.6　平台的视频录像回放

10.5.2　视图无缝融合功能

平台在统一的界面上实现了对视频和图像两类监控设备的管理操作（如添加、删除、编辑），如 DVR 等视频类监控设备和智能卡口等图像类监控设备，可在同一管理终端，同时查看两类设备的运行状态；依据系统设置的关联关系，可通过图像搜索对应的视频录像；在提示现场异常情况后，可控制云台转动，多角度查看现场状态。

平台提供图中画和图表播放模式，可根据地图查找通道，直接观看地图中的通道视频并进行通道操作。通过直观的图表式地图播放，快速切换到监控点，支持打开多个视频窗口。在电子地图中支持矢量地图，形象地标识出摄像机的地理位置，无须对地图进行切换，只要通过地图缩放，就可以寻找到所有的摄像机位置。

如图 10.7 所示，平台提供分层电子地图。在产生报警或故障时，可通过电子地图准确显示事发位置。支持多级树状结构，具有图层跳转功能。支持矢量地图，可自由缩放。

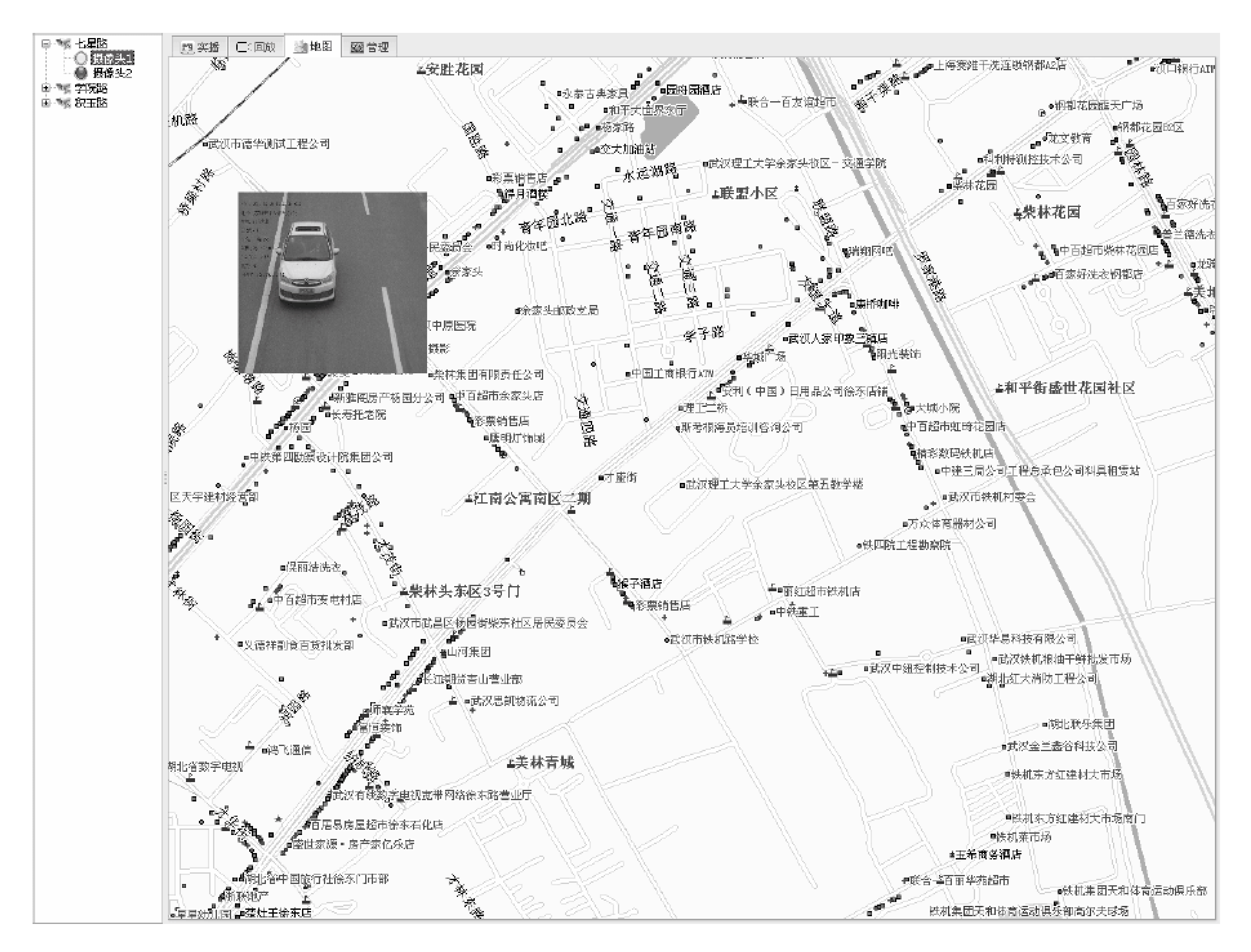


图 10.7　平台的视图融合

10.5.3　大规模人脸等目标监测

布撤控是海量视频管控平台的重要功能，可对重要目标（如某人或车）设置重点关注，当监控视频中出现此人或车时，系统智能检测和发布警情，并在数据库中记录报警信息，支持分类查询。

可根据时间、地点、车牌等信息，对目标车辆进行全方位布控；对车牌号码记录不全的车辆，支持通配符模糊布控；多条件检索布控状态，可以根据布控属性查找已经布控的记录。

布控分等级，优先级别高的布控项目优先提示。在系统繁忙时，能够保证重点关注和重要信息。

支持手动和自动撤控，支持布控信息的批量导入/导出。

支持视频预案功能，依据具体需求设计监控预案，可直接控制到各监控点的监控时间和预置位，为其提供更为直观的功能显示和屏幕操作。

支持报警预案配置功能，提供多种报警联动策略（如声音、指示灯、视频切换、视频放大、视频上墙、云台预置点、视频预案等），可对不同报警设置预规划响应（如时

间、场景等），支持可疑情况防范功能。

10.5.4　异常行为检测

异常行为主要包括暴力行为和可疑行为，异常行为检测涉及计算机视觉、图像处理、模式识别和人工智能等多个学科，它采用视觉机器学习方法，分析监控场景的视频数据，提取异常行为的显著性和稳健性特征，判别场景中是否存在异常活动。

异常行为检测包括用户交互模块、视频转换模块、行为检测模块、数据存储与显示模块等，其核心是行为检测模块。

10.5.5　海量视频摘要

如图 10.8 所示，针对卡口、电子警察等设备传输到平台的视频图像，通过视频分析与计算方法，浓缩产生视频图像的属性和语义信息。

视频图像的有效信息可同步显示在监控窗口下方，如时间、地点、车牌号码、归属地等。

视频摘要信息可直接关联录像，呈现事件发生的前因后果，可单独放大以查看细节。



图 10.8　视频摘要图像

10.5.6　高清卡口车辆信息搜索

高清卡口车辆信息搜索主要包括车流量查询和违法事件搜索。

1. 车流量查询

平台对卡口车流量进行自动统计，可设置搜索条件或查询要求。

图 10.9 显示了平台自动绘制的柱状图，可直观显示统计结果；可对不同车型进行分类统计；可按照日/周/月/年统计，生成报表。

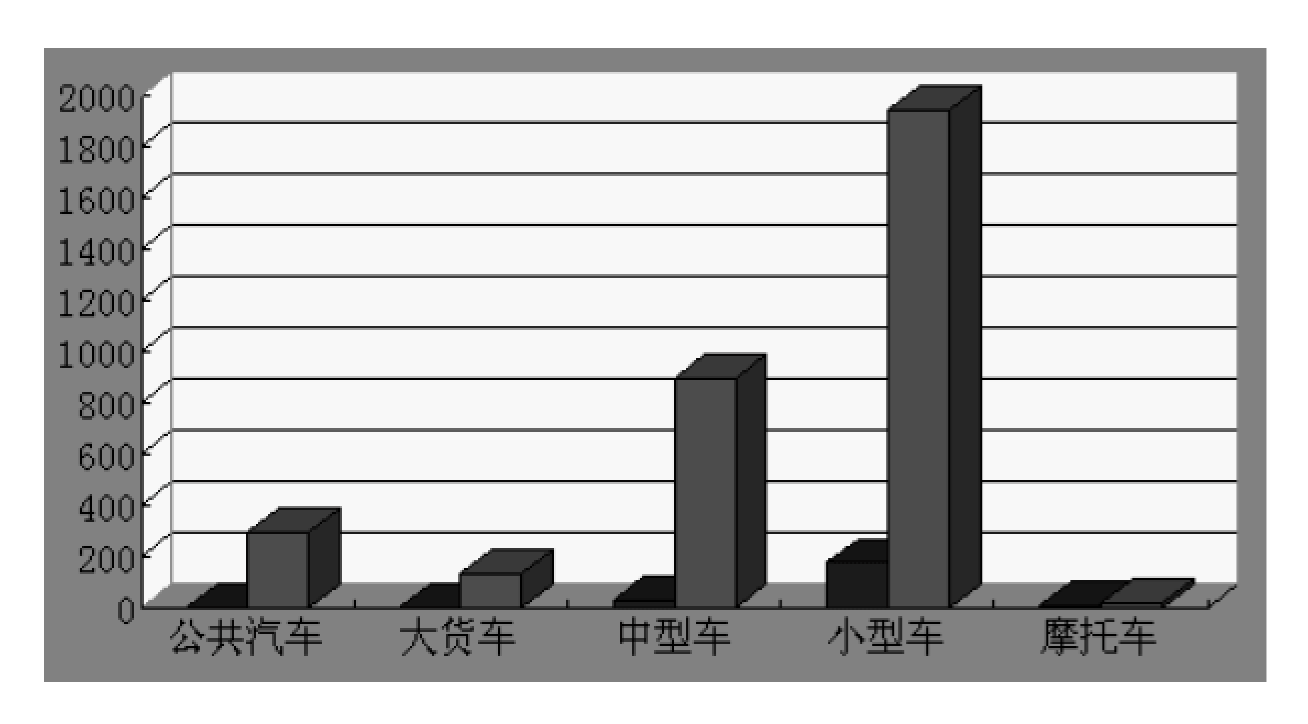


图 10.9　车流量统计

2. 违法事件搜索

如图 10.10 所示，可按违章事件存放车辆图像以及相关视频，实现录像信息按事件分类，通过违章事件搜索所有违章图像，关联播放相关视频。

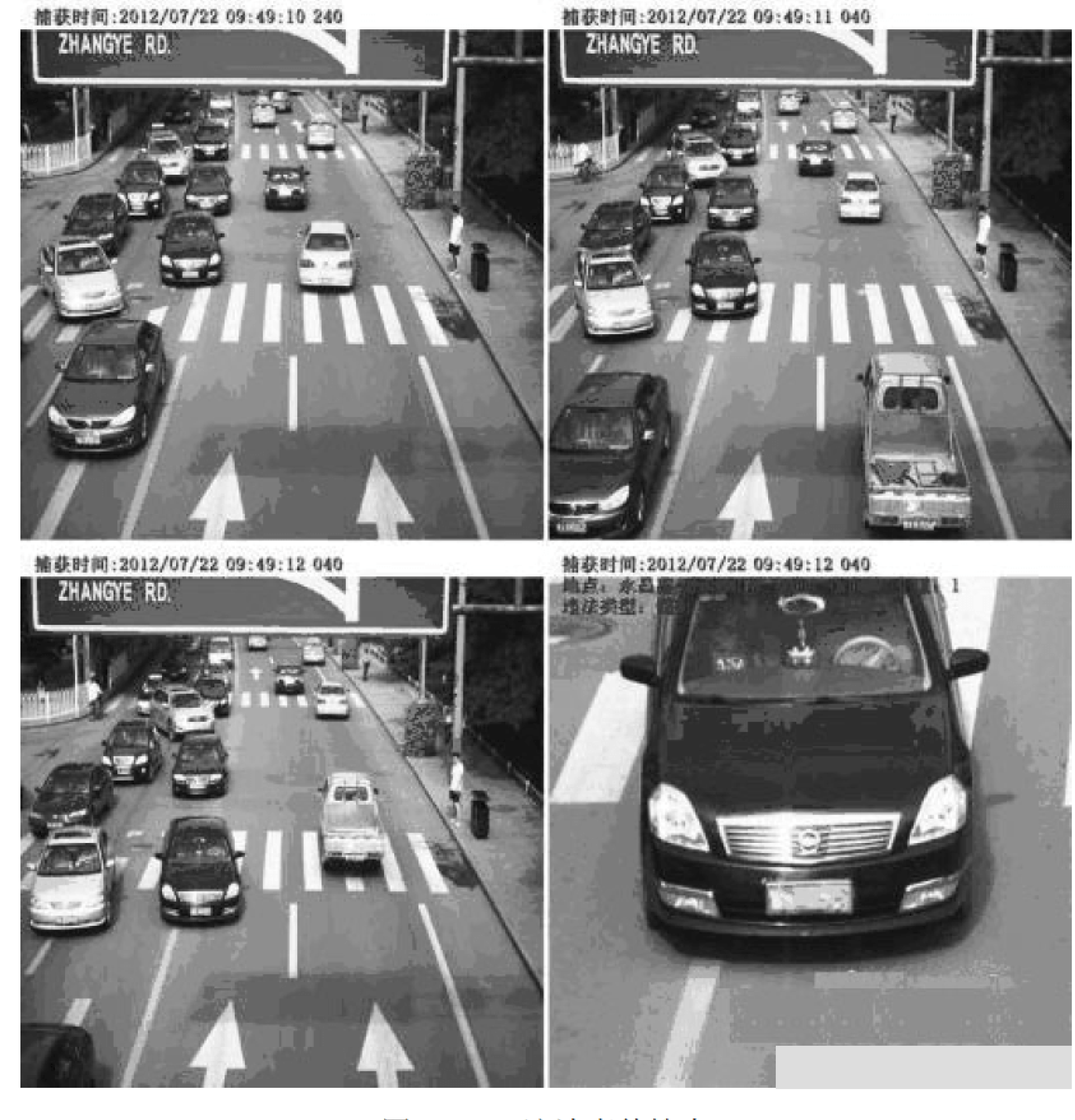


图 10.10　违法事件搜索

支持远程 DVR、点播服务器中视频文件的搜索，可按照视频通道、录像类型、存放位置、车牌、车标、车型、关键字、时间等条件进行搜索，监控中心能按地域、图像通道、日期和时间对前端设备进行视频文件搜索。

10.6 平台应用

如图 10.11 所示，海量视频管控平台的主要应用如下。

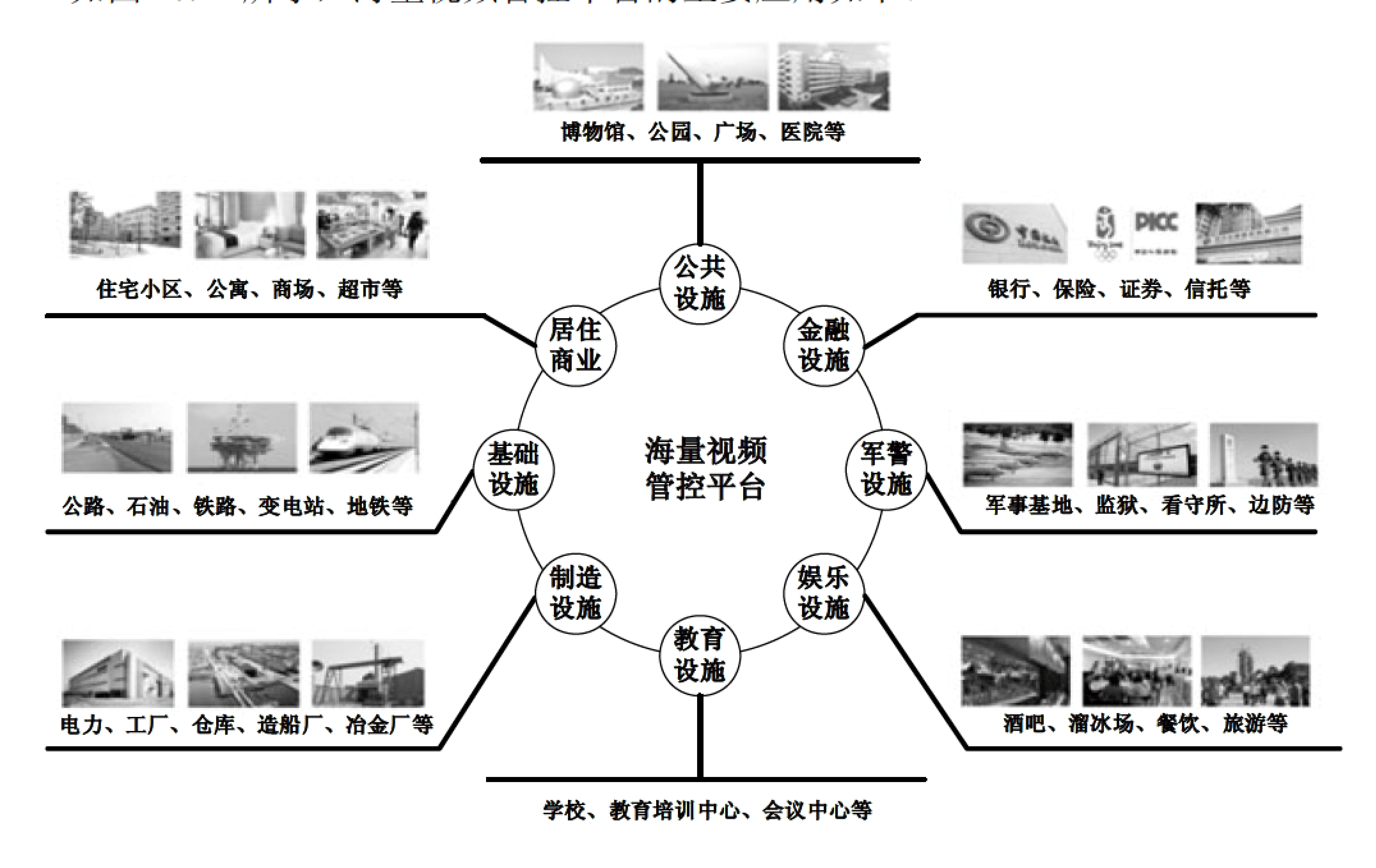


图 10.11 海量视频管控平台的应用

可见，海量视频监控平台目前已广泛应用于公共设施、金融设施、军警设施、娱乐设施、教育设施、制作设施、基础设置和居民商业领域。随着技术的革新和完善，其应用领域还将进一步拓展。